



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ

Ανάλυση δεδομένων αγώνων μπάσκετ

Πτυχιακή εργασία

ΣΠΥΡΟΣ ΚΑΝΕΛΛΟΠΟΥΛΟΣ

Αθήνα, Φεβρουάριος 2022



HAROKOPIO UNIVERSITY

SCHOOL OF DIGITAL TECHNOLOGY
DEPARTMENT OF INFORMATICS AND
TELEMATICS

Data Analytics & Trajectory Mining in Basketball Games

Bachelor thesis

SPYROS KANELLOPOULOS

Athens, February 2022



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ

Βαρβακάρη Μαλβίνα
Επίκουρος Καθηγητής,
Τμήμα Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο

Δημητρακόπουλος Γεώργιος
Επίκουρος Καθηγητής,
Τμήμα Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο

Καραγιώργου Σοφία
Επισκέπτης Λέκτορας,
Τμήμα Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο

Ο Σπύρος Κανελλόπουλος

δηλώνω υπεύθυνα ότι:

- 1 Είμαι ο κάτοχος των πνευματικών δικαιωμάτων της πρωτότυπης αυτής εργασίας και από όσο γνωρίζω η εργασία μου δε συκοφαντεί πρόσωπα, ούτε προσβάλει τα πνευματικά δικαιώματα τρίτων.
- 2 Αποδέχομαι ότι η ΒΚΠ μπορεί, χωρίς να αλλάξει το περιεχόμενο της εργασίας μου, να τη διαθέσει σε ηλεκτρονική μορφή μέσα από τη ψηφιακή Βιβλιοθήκη της, να την αντιγράψει σε οποιοδήποτε μέσο ή/και σε οποιοδήποτε μορφότυπο καθώς και να κρατά περισσότερα από ένα αντίγραφα για λόγους συντήρησης και ασφάλειας.

Σελίδα για Αφιέρωση (προαιρετικά)

Σελίδα για γνωμικό (προαιρετικά)

Ευχαριστίες

Ολοκληρώνοντας τη πτυχιακή μου εργασία, θα ήθελα να ευχαριστήσω όλους όσους βοήθησαν σε όλα τα στάδια ανάπτυξής της. Θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτρια κ. Καραγιώργου Σοφία για την εξαιρετική συνεργασία καθώς και τις εύστοχες υποδείξεις καθ' όλη τη διάρκεια της εκπόνησης της εργασίας. Η προθυμία και η διάθεσής της για αυτήν την ερευνητική εργασία με βοήθησαν να μελετήσω και να αναπτύξω ένα αντικείμενο που προσωπικά με ενθουσιάζει.

Πίνακας Περιεχομένων

1 Εισαγωγή	16
1.1 Εξόρυξη δεδομένων τροχιάς	16
1.2 Αντικείμενο πτυχιακής.....	19
1.2.1 Ερευνητικό πρόβλημα	20
1.2.2 Συνεισφορά.....	21
1.2.3 Περιορισμοί	22
1.3 Οργάνωση κειμένου	23
2 Σχετικές δραστηριότητες και εργασίες	24
2.1 Οπτικοποίηση αθλητικών συστημάτων μέσω δεδομένων εξόρυξης τροχιάς	24
2.2 Εξαγωγή και βαθμονόμηση θερμικών χαρτών από καταγραφές αθλητικών εφαρμογών	26
2.3 Προσδιορισμός και κατανόηση αστικών αθλητικών περιοχών χρησιμοποιώντας το Nokia Sports Tracker	27
2.4 Διερεύνηση πολυπλοκότητας της ανθρώπινης κίνησης δια μέσου δεδομένων τροχιάς.....	28
3 Τεχνολογικό υπόβαθρο.....	30
3.1 Στάδιο δημιουργίας ψηφιακού περιβάλλοντος.....	31
3.2 Στάδιο οργάνωσης αρχείων εφαρμογής	32
3.3 Στάδιο συλλογής αθλητικών στατιστικών δεδομένων.....	35
3.4 Στάδιο αποθήκευσης αθλητικών στατιστικών δεδομένων.....	53
3.5 Στάδιο επεξεργασίας αθλητικών στατιστικών δεδομένων	57
3.5 Οπτικοποίηση αποτελεσμάτων	60
4 Πειραματικά αποτελέσματα	62
4.1 Παρουσίαση αποτελεσμάτων.....	63
4.2 Οπτικοποίηση συμπερασμάτων.....	64

5 Επίλογος	74
5.1 Σύνοψη και συμπεράσματα	74
5.2 Μελλοντικές επεκτάσεις.....	77
Βιβλιογραφία	78

Περίληψη

Ένα σημαντικό καθήκον για τη βελτίωση της απόδοσης μιας αθλητικής ομάδας είναι η ανάλυση προηγούμενων αγώνων, από διάφορες οπτικές. Υπάρχουν διάφοροι τύποι μεθόδων ανάλυσης αθλητικής στρατηγικής που αντιμετωπίζονται στα παιχνίδια μπάλας, ένας από αυτούς είναι στατιστική ανάλυση. Στη στατιστική ανάλυση, η βαθμολόγηση πληροφοριών αναλύεται με στατιστικές μεθόδους, έτσι ώστε να ληφθεί η καλύτερη δυνατή απόφαση με τα παρών δεδομένα.

Η ανάλυση στρατηγικής μπορεί να κάνει προβλέψεις για μελλοντικά παιχνίδια και να υποστηρίξει αποφάσεις στρατηγικής. Αν και μπορεί να ληφθεί ποικιλία δεδομένων για το παιχνίδι, είναι δύσκολο να προσδιοριστούν οι πληροφορίες που θέλουν οι προπονητές από τα δεδομένα. Στην στατιστική ανάλυση, πληροφορίες στο σκοράρισμα όπως χρόνος σκοραρίσματος, τοποθεσία και ποσοστό ευστοχίας βοηθούν στην πρόβλεψη του επόμενου παιχνιδιού. Με αυτόν τον τρόπο, όταν παρουσιάζονται παρόμοιες καταστάσεις μπορούμε να προβλέψουμε σε κάποιο ποσοστό το αποτέλεσμα ή να απομονώσουμε ένα συγκεκριμένο χαρακτηριστικό(απόδοση παίκτη).

Η παρούσα εργασία στοχεύει στην απεικόνιση των προαναφερθέντων στατιστικών αλλά και σε έναν μελλοντικό τρόπο αξιοποίησης τους. Η αξιοποίηση των στατιστικών γίνεται μέσω γραφημάτων έτσι ώστε να αποφεύγεται η επανάληψή τους αλλά και για την σαφέστερη οπτικοποίηση των σχέσεων των οντοτήτων. Επιπλέον, δίνεται η δυνατότητα τροποποίησης των στατιστικών είτε προσθέτοντας είτε αφαιρώντας παραμέτρους, προσφέροντας έτσι την αλληλεπίδραση με τον χρήστη.

Συνοψίζοντας, τα γραφήματα δίνουν μια σαφή εικόνα της υπεροχής μιας ομάδας ή ενός παίκτη. Με την παρούσα δομή, γίνεται εύκολα αντιληπτή η έφεση μιας ομάδας στο σκοράρισμα ή αντίστοιχα στην άμυνα. Ο κάθε παίκτης περιγράφεται από την ευστοχία του αλλά και από την συνολική συνεισφορά του στον αγώνα, προσδιορίζοντας έτσι το συνολικό φάσμα των δυνατοτήτων του. Με τα

παραπάνω δεδομένα δίνεται η δυνατότητα δημιουργίας ομάδας με βάση τα χαρακτηριστικά των παικτών τα οποία ανταποκρίνονται σε πραγματικά στατιστικά.

Λέξεις κλειδιά: Δεδομένα εξόρυξης τροχιάς, εφαρμογή Crud, πλαίσιο Django, Python, HTML, JavaScript, διαγράμματα, στατιστικά αθλητισμού

Abstract

An important task in improving the performance of a sports team is to analyze previous games from different perspectives. There are several types of sports strategy analysis methods used in ball games, one of which is statistical analysis. In statistical analysis, the rating of information is analyzed by statistical methods, to make the best possible decision with the present data.

Strategy analysis can make predictions for future games and support strategy decisions. Although a variety of game data can be obtained, it is difficult to identify the information that coaches want from the data. In statistical analysis, scoring information such as scoring time, location and accuracy rate help predict the next game. In this way, when similar situations occur, we can predict the result to some extent or isolate a specific feature (player performance).

The present work aims at the depiction of the above-mentioned statistics but also at a future way of their utilization. The utilization of statistics is done through graphs to avoid their repetition but also for a clearer visualization of the relationships of the entities. In addition, it is possible to modify the statistics either by adding or removing parameters, thus offering interaction with the user.

In summary, the graphs give a clear picture of the superiority of a team or a player. With the present structure, it is easy to understand the appeal of a team in scoring or respectively in defense. Each player is described by his aptness but also by his overall contribution to the match, thus determining the overall range of his capabilities. With the above data it is possible to create a team based on the characteristics of the players which correspond to real statistics.

Keywords:

Track mining data, Crud application, Django framework, Python, HTML, JavaScript, charts, sports statistics

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Αναφορά αθλητή Γραπτό έντυπο στο οποίο απεικονίζεται η αγωνιστική απόδοση του αναφερόμενου αθλητή. Χρησιμοποιούνται για στατιστικούς λόγους. ..	19
Εικόνα 2: Στιγμιότυπο βάσης δεδομένων	24
Εικόνα 3: Οπτικοποίηση συστήματος ποδοσφαίρου μέσω διαγράμματος τροχιάς.....	26
Εικόνα 4: Pycharm λογότυπο	35
Εικόνα 5: Στιγμιότυπο Mysql βάσης δεδομένων.....	57
Εικόνα 6: Εισαγωγή μέσω SQL queries	57
Εικόνα 7: Προεπιλεγμένη φόρμα συμπλήρωσης στοιχείων	59
Εικόνα 8: Μενού UI για την κατηγοριοποίηση	65

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1: Χαρακτηριστικά Κλάσης Team	37
Πίνακας 2: Χαρακτηριστικά Κλάσης Player	39
Πίνακας 3: Χαρακτηριστικά Κλάσης TeamStats	42
Πίνακας 4: Χαρακτηριστικά Κλάσης PlayerStats	46
Πίνακας 5: Χαρακτηριστικά Κλάσης AverageTeamStats	50
Πίνακας 6: Χαρακτηριστικά Κλάσης AveragePlayerStats.....	53

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1: Σχεδιάγραμμα για την καλύτερη απεικόνιση της django δομής	34
Σχήμα 2: Τελικό σχεδιάγραμμα django δομής	35

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

Διάγραμμα 1: Κορυφαίες ομάδες σε ποσοστό τριπόντων	67
Διάγραμμα 2: Διευρυμένο διάγραμμα που απεικονίζει εύστοχες και συνολικές προσπάθειες.....	68
Διάγραμμα 3: Διάγραμμα μπάρας που απεικονίζει τις τρεις κορυφαίες ομάδες σε αμυντικών διεκδικήσεων	69
Διάγραμμα 4: Διάγραμμα μπάρας που απεικονίζει τις τρεις κορυφαίες ομάδες σε πόντους.....	70
Διάγραμμα 5: Διάγραμμα πίτας που απεικονίζει τις τρεις κορυφαίες ομάδες σε εύστοχες βολές.....	71

Διάγραμμα 6: Διάγραμμα σφαίρας που απεικονίζει τους πέντε κορυφαίους παίκτες σε κλειψίματα.	72
Διάγραμμα 7: Διάγραμμα σφαίρας που απεικονίζει τους πέντε κορυφαίους παίκτες σε τελικές πάσες.	73

ΠΙΝΑΚΑΣ ΑΠΟΣΠΑΣΜΑΤΩΝ ΚΩΔΙΚΑ

Απόσπασμα κώδικα 1: Ορισμός κλάσης PlayerStats	46
Απόσπασμα κώδικα 2: Credentials βάσης δεδομένων	55
Απόσπασμα κώδικα 3: Αρχείο views.py	60
Απόσπασμα κώδικα 4: Ρυθμίσεις settings.py για την χρήση templates	61
Απόσπασμα κώδικα 5: html κώδικας με.....	62

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

CRUD	Create Read Update Delete
NBA	National Basketball Association
API	Application Programming Interface
JS	JavaScript
3PP	Three Point Percentage
2PP	Two Point Percentage
DB	Database
DoB	Date of Birth

1 *Εισαγωγή*

Στο κεφάλαιο αυτό γίνεται μια εισαγωγή στη θεματική ενότητα των τεχνικών εξόρυξης δεδομένων και συγκεκριμένα μέσω της αξιοποίησης δεδομένων τροχιάς στο ευρύτερο πλαίσιο της τεχνολογίας. Παρουσιάζεται αναφορικά η εξέλιξη τους καθώς και ορισμένα από τα πεδία στα οποία έχουν άμεση εφαρμογή. Επίσης, αναλύεται ο σκοπός της παρούσας εργασίας και δίνεται μια αποτύπωση της δομής που ακολουθείται στη συνέχεια αυτής.

1.1 Εξόρυξη δεδομένων τροχιάς

Η επιστήμη της εξόρυξης δεδομένων (Data Mining) είναι μία από τις κυρίαρχες ερευνητικές κατευθύνσεις και αποτελεί αντικείμενο πολυάριθμων μελετών τα τελευταία έτη, διότι παρουσιάζει μεγάλο εύρος και ποικιλομορφία στα πεδία εφαρμογών και υπηρεσιών που υποστηρίζει.

Εξόρυξη δεδομένων(ή ανακάλυψη γνώσης από βάσεις δεδομένων) είναι η εξεύρεση μιας ενδιαφέρουσας, αυτονόητης, μη προφανούς και πιθανόν χρήσιμης πληροφορίας ή προτύπων (patterns) από μεγάλες βάσεις δεδομένων με χρήση αλγορίθμων ομαδοποίησης ή κατηγοριοποίησης και των αρχών της στατιστικής, της τεχνητής νοημοσύνης (Artificial Intelligence), της μηχανικής μάθησης (Machine Learning) και των συστημάτων βάσεων δεδομένων. Στόχος της εξόρυξης δεδομένων είναι η πληροφορία που θα εξαχθεί και τα πρότυπα που θα προκύψουν να έχουν

δομή κατανοητή προς τον άνθρωπο έτσι ώστε να τον βοηθήσουν να πάρει τις κατάλληλες αποφάσεις.

Ένας από τους τομείς που έχει άμεση εφαρμογή η επιστήμη της εξόρυξης πληροφορίας (data mining) είναι τα δεδομένα τροχιάς. Οι χωρικές πληροφορίες κινούμενων αντικειμένων, όπως οι αθλητές, που παρέχονται σε συνδυασμό με τα ποσοστά επιτυχίας τους αποτελούν την κύρια αιτία της έρευνας αυτής. Τα δεδομένα αυτά, παρέχονται καθόλη την διάρκεια της αγωνιστικής περιόδου με την βοήθεια του ανθρώπινου παράγοντα. Η εμπλοκή του παράγοντα αυτού ωστόσο, καθιστά τα δεδομένα αυτά επιρρεπή στα λάθη, γεγονός που μπορεί να αποφευχθεί με την συμβολή της τεχνολογίας. Η μορφή των πληροφοριών παρουσιάζεται με σχεδιάγραμμα της παρούσας τοποθεσίας του αθλητή και τον αριθμό προσπαθειών που έχει ευστοχήσει ή αστοχήσει από την τοποθεσία αυτή (βλέπε εικόνα 1). Επιπλέον, αναφέρονται και χαρακτηριστικά όπως τελικές πάσες, διεκδικήσεις, λάθη και κοψίματα (assists, rebounds, turnovers, blocks) που αναδεικνύουν την συμβολή του αθλητή σε άμυνα και επίθεση. Τα ακατέργαστα αυτά δεδομένα, είναι δύσχρηστα με την πρώτη όψη, ωστόσο προκύπτουν χρήσιμα συμπεράσματα όπως ποσοστά ευστοχίας επί τοις εκατό σε κάθε τοποθεσία αλλά και γενικότερη συμμετοχή στην επίθεση λχ assist. Οι πληροφορίες αυτές αποθηκεύονται, με στόχο την ανάλυση της ατομικής διακύμανσης του κάθε παίκτη σε πλήθος αγώνων αλλά και την συνολική του συνεισφορά σε ομαδικό επίπεδο.

Individual Player Scouting Report

Name JONES JR Bryce Tyler Number 32 Position PG
 Height 6.3 Weight 170 Build Thick Dominant Hand RIGHT



X - MISSED SHOTS

O - MADE SHOTS

- ASSISTS ### ///
- REBOUNDS ###
- TURNOVERS ///
- BLOCKS

	STRONG	FAIR	WEEK
Outside Shooting		X	
Mid-Range Shooting		X	
Inside Shooting	X		
Passing Ability	X		
Ball Handling Ability	X		
Off Hand Dribbling Ability		X	
Decision Making	X		
One on One Offense	X		
Dribble Penetration	X		
Quickness		X	
Offensive Rebounding			X
Defensive Rebounding			X
Box Out		X	
On Ball Defense		X	
Off Ball Defense		X	
Off Ball Movement		X	
Leadership	X		
Aggressiveness		X	
Free Throws Shooting	X (5-4)		

Comments: PLAYER IS DOMINATING THE GAME WITH THE BALL IN HIS HANDS.
- STRONG PICK AND ROLL PLAYER

Εικόνα 1: Αναφορά αθλητή

Γραπτό έντυπο στο οποίο απεικονίζεται η αγωνιστική απόδοση του αναφερόμενου αθλητή.

Χρησιμοποιούνται για στατιστικούς λόγους.

Μερικά από τα ερευνητικά ζητήματα που σχετίζονται με το πεδίο αυτό είναι η αποτελεσματική απεικόνιση των σύνθετων δεδομένων και των επαναλαμβανόμενων προτύπων που εμφανίζονται σε αυτά. Ειδικότερα, είναι εμφανής η δυσκολία αναγνώρισης των προτύπων αυτών, αφού εμπλέκονται αστάθμητοι παράγοντες όπως η κόπωση εφόσον μιλάμε για ένα άθλημα. Συνεπώς, δεν είναι ακριβές να παρομοιάσουμε δυο καταστάσεις που συνέβησαν στην αρχή και στο τέλος ενός αγώνα. Για την καλύτερη απεικόνιση τέτοιων δομών στρεφόμαστε σε περισσότερο σύνθετα γραφήματα όπως ποσοστά ευστοχίας έπειτα από συγκεκριμένο χρόνο, στα οποία ωστόσο δεν γίνεται αναφορά στο πλαίσιο αυτής της έρευνας.

Ευρύτερα μιλώντας, η στατιστική απεικόνιση στα αθλήματα δίχως την γενικότερη εικόνα είναι ελλιπής όταν εμπλέκεται ο ανθρώπινος παράγοντας. Μια ατομική ανάλυση απόδοσης, στα ομαδικά αθλήματα μπορεί να επηρεαστεί από διάφορους συντελεστές, όπως είναι μια επιδέξια ομάδα. Σε αυτήν την περίπτωση, ο προπονητής στοχεύει στην ανάδειξη και στην αύξηση των προσωπικών αυτών στατιστικών. Απεναντίας, βρίσκοντας αντιμέτωπος με έναν ικανό αντίπαλο, τα ατομικά στατιστικά αυτά θα παρουσιάσουν μια μείωση δίχως να φαίνεται η αιτία.

Για τους παραπάνω λόγους είναι δύσκολη η αποκλειστική ύπαρξη των στατιστικών δίχως τον ανθρώπινο παράγοντα να τις διαχειρίζεται.

Με δεδομένη την εξέλιξη των στατιστικών μεθόδων και γραφημάτων παρατηρούμε μια καλύτερη ανάλυση των ατομικών επιδόσεων των αθλητών. Για την ακρίβεια, στα ομαδικά αθλήματα παρατηρούμε νέους τρόπους απεικόνισης της συμβολής τους, λχ ρυθμός καλαθιών(NetRtg). Ο ρυθμός καλαθιών αποτελεί πόσους πόντους έβαλε μια ομάδα αφαιρώντας τον αριθμό που δέχθηκε, σε μια συγκεκριμένη χρονική περίοδο όσο αυτός ο παίκτης βρισκόταν στο παρκέ. ($\text{NetRtg} = \text{OffRtg} - \text{DefRtg}$ όπου $\text{OffRtg} = \text{points scored per 100 possessions}$ και $\text{DefRtg} = \text{points allowed per 100 possessions}$). Με την ύπαρξη της παρούσας ένδειξης, γίνεται αντιληπτή μέσω γενικότερων στατιστικών, πόσο σημαντική είναι η συμμετοχή ενός αθλητή δίχως να γίνεται αναφορά στα προσωπικά του στατιστικά.

1.2 Αντικείμενο πτυχιακής

Αντικείμενο της παρούσας μελέτης είναι να αναλύσει τα στατιστικά στοιχεία που προκύπτουν από την απόδοση αθλητών μπάσκετ και των ομάδων τους, τις προοπτικές εξέλιξης και χρήσης αλγοριθμικών τεχνικών για την αυτοματοποιημένη δημιουργία εξειδικευμένων στατιστικών. Λαμβάνοντας υπόψιν τη δυσκολία συγκέντρωσης, ομαδοποίησης και διαμόρφωσης των αθλητικών επιδόσεων που απαιτούνται για την κατασκευή μοντέλων(αθλητών -ομάδων) σε συνδυασμό με την ύπαρξη ανάγκης παραγωγής ποιοτικών αποτελεσμάτων, καθίσταται εμφανές ότι αναμένεται λεπτομερής ανάλυση των δεδομένων πριν τη χρήση τους. Απαιτείται, επομένως, μια εκτενής και σε πολλά επίπεδα μελέτη της πληθώρας των δεδομένων η οποία θα οδηγήσει στην εύρεση του τρόπου της κατάλληλης επεξεργασίας τους.

Επιπρόσθετα, με τη χρήση των αποτελεσμάτων από τη διερεύνηση της επίδοσης των αθλητών και την γραφική απεικόνισή τους, επιδιώκεται να προσδιοριστούν οι παράγοντες που οδηγούν στην επιτυχία του/της. Επιπλέον στοχεύει στην πρόβλεψη επίδοσεις αθλητών-ομάδων κάτω από συγκεκριμένες συνθήκες.

Τα προαναφερθέντα παρατίθενται με σκοπό να διαμορφώσουν ένα γνωστικό πεδίο πάνω στο οποίο μπορούν να στηριχθούν τεκμηριωμένα τρόποι και κριτήρια στην σχεδίαση των αθλητικών ομάδων μπάσκετ.

1.2.1 Ερευνητικό πρόβλημα

Κυρίαρχο πρόβλημα θεωρείται η ύπαρξη και η συντήρηση μιας εχεμύθου βάσης δεδομένων. Αναλυτικότερα, στο κορυφαίο αγωνιστικό επίπεδο (National Basketball Association), είναι αμφίβολη η ύπαρξη ακριβών στατιστικών, με δεδομένο πως η πληροφορία θα είναι προσβάσιμη σε όλους, συνεπώς και στις αντίπαλες ομάδες. Για τον παραπάνω λόγο είναι αντίξοη η ύπαρξη τέτοιων στατιστικών αφού θα βοηθήσει το έργο της ομάδας αλλά ταυτόχρονα θα διευκολύνει και την αντιμετώπισή της.

Το δεύτερο ερευνητικό πρόβλημα που παρουσιάζεται στην παρούσα έρευνα αφορά τον ανθρώπινο παράγοντα. Πιο συγκεκριμένα, αντικείμενο της μελέτης αυτής αποτελεί η διαλογή συγκεκριμένων μεταβλητών από ένα πλήθος στατιστικών παραμέτρων. Με την βοήθεια των μεταβλητών αυτών, ο ανθρώπινος παράγοντας (προπονητική ομάδα) ορίζει την πορεία δράσης της ομάδας. Ωστόσο, η αρμόδια προπονητική ομάδα αποτελεί αστάθμητο παράγοντα καθώς παράγει μη σταθερά αποτελέσματα, φαινόμενο που θα αναλυθεί παρακάτω.

1.2.2 Συνεισφορά

Η εν λόγω έρευνα αποτελεί μια διαδικτυακή εφαρμογή αποθήκευσης (Create, Read, Update, Delete - CRUD web app) συσσωρεύει στατιστικά στοιχεία ομάδων και παικτών. Οι εφαρμογές αποθήκευσης διακρίνονται από τα εξής χαρακτηριστικά: δημιουργία, ανάγνωση, ενημέρωση, διαγραφή (Create, Read, Update, Delete). Κάθε διαδικτυακή εφαρμογή οφείλει να καλύπτει τα παραπάνω χαρακτηριστικά για την ομαλή λειτουργία της. Στην παρούσα μελέτη, η επιλογή της παραπάνω δομής προσφέρει τα ακόλουθα γνωρίσματα:

1) Διαδραστικότητα του χρήστη. Οποτεδήποτε δημοσιεύεται μια έρευνα σε μια ιστοσελίδα που περιέχει στατιστικά στοιχεία ο επισκέπτης δεν μπορεί να συμμετάσχει στην έρευνα αυτή παρά μόνο να αναγνώσει τα δεδομένα που του παρέχονται. Απεναντίας στις διαδικτυακές εφαρμογές, ο χρήστης μπορεί εκτός από

την ανάγνωση στοιχείων, να αλληλεπιδράσει με αυτά. Στην εν λόγω εφαρμογή, ο εξουσιοδοτημένος, από τον διαχειριστή, χρήστης έχει την δυνατότητα να πραγματοποιήσει οποιαδήποτε αλλαγή επιθυμεί στην ομάδα του. Αυτό έχει ως επακόλουθο, η προπονητική ομάδα να πραγματοποιήσει όσες μεταγραφές θέλει μέχρις ότου να συναντήσει τα ποσοστά που εκείνη θέλει.

2) Ενσωμάτωση (integration). Με τον όρο ενσωμάτωση, εννοούμε την ένωση πολλών διαφορετικών στοιχείων, με στόχο την δημιουργία μιας περισσότερο σύνθετης δομής. Ο σχεδιασμός της ιστοσελίδας επιδρά αρνητικά στην ύπαρξη μιας πολυδιάστατης μορφής. Εν αντιθέσει, η πολυσύνθετη λειτουργικότητα της εφαρμογής παρέχει την δυνατότητα επικοινωνίας με εξωτερικά συστήματα.

3) Καλύτερη οργάνωση δεδομένων. Εφόσον είναι μια διαδικτυακή εφαρμογή, η οργάνωση, πρόσβαση και επεξεργασία των δεδομένων είναι πιο εύκολη. Η πλειοψηφία των ιστοσελίδων με στατιστικά στοιχεία αποτελούν μια στατική δομή. Στην παρούσα εφαρμογή, η διαχείριση των δεδομένων γίνεται δυναμικά και πιο εύχρηστα. Πιο συγκεκριμένα, προγραμματιστικά, χρησιμοποιήθηκαν κλάσεις χαρακτηριστικό που διευκολύνει την οργάνωση και επαναχρησιμοποίηση των δεδομένων αυτών δίχως επανάληψη κώδικα (λιγότερος φόρτος για Η/Υ).

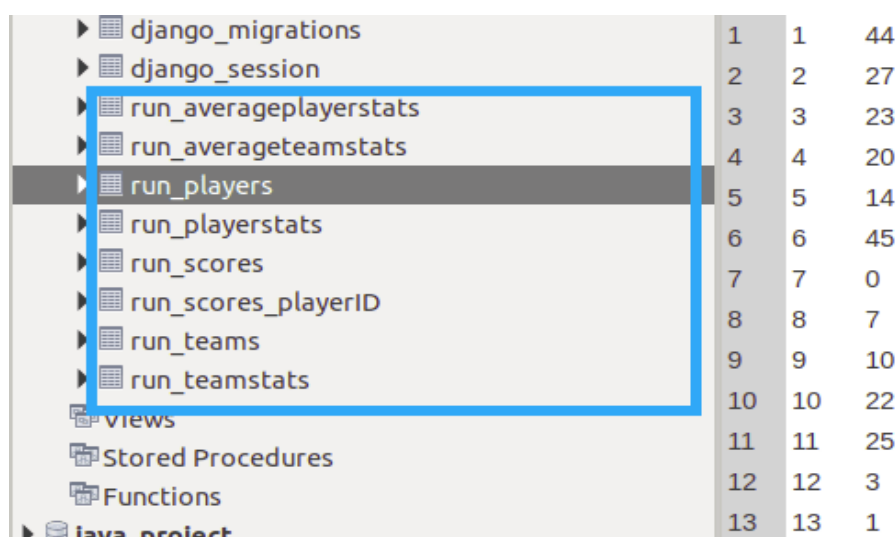
4) Δυναμική απεικόνιση δεδομένων. Ευρύτερα, οι ιστοσελίδες που παρέχουν στατιστικές πληροφορίες θεωρούν πως το σύνολο των επισκεπτών τους διαθέτουν εξειδικευμένες αθλητικές γνώσεις. Ωστόσο, δεν αποτελούν όλοι οι επισκέπτες μέλη προπονητικής ομάδας, συνεπώς να είναι εξοικειωμένοι αθλητικές ορολογίες. Στην παρούσα εφαρμογή, έχουν αξιοποιηθεί πλήθος διαγραμμάτων (bullet diagram, bar diagram, pie diagram), τα οποία παρέχουν πληροφορίες αντιληπτές στον απλό χρήστη αλλά ταυτόχρονα και εξειδικευμένες για τις προπονητικές ομάδες.

1.2.3 Περιορισμοί

Κυριότερος ανασταλτικός παράγοντας αποτελεί η συλλογή δεδομένων. Στις ερευνητικές εργασίες, όσο πιο εύκολα μεταφέρονται οι πληροφορίες, τόσο πιο άμεσα γίνεται η επεξεργασία τους και με μικρότερο περιθώριο λάθους. Συνεπώς, έχουμε ταχύτερη και περισσότερο αξιόπιστη παραγωγή αποτελεσμάτων.

Επιπρόσθετα, εμπόδιο αποτελεί στην έρευνα, η χρήση πολλαπλών πηγών που παρουσιάζουν ανόμοια αποτελέσματα. Το γεγονός αυτό, δημιουργεί αναξιοπιστία ως προς την εξαγωγή ασφαλών συμπερασμάτων. Για τον ίδιο λόγο αντίστοιχα, δεν μπορεί να γίνει εις βάθος έρευνα αφού δεν είμαστε σίγουροι για την αξιοπιστία των αποτελεσμάτων, συνεπώς και για το πόρισμα που θα δημιουργηθεί.

Στην παρούσα μελέτη, χρησιμοποιήθηκαν στατιστικά δεδομένα από τον ιστότοπο basketball reference (<https://www.basketball-reference.com/>). Οι πληροφορίες αυτές αντιγράφηκαν μέσω sql queries, τοπικά στην βάση, στα πλαίσια της μελέτης. Ως αποτέλεσμα, η αντιγραφή των δεδομένων αυτών είναι επιρρεπής σε λάθη, αφού δεν μιλάμε για μια αυτόματη διαδικασία αλλά για μια χειροκίνητη.



1	1	44
2	2	27
3	3	23
4	4	20
5	5	14
6	6	45
7	7	0
8	8	7
9	9	10
10	10	22
11	11	25
12	12	3
13	13	1

Εικόνα 2: Στιγμιότυπο βάσης δεδομένων

1.3 Οργάνωση κειμένου

Η συνέχεια της παρούσας μελέτης οργανώνεται σύμφωνα με την περιγραφή που ακολουθεί. Στο Κεφάλαιο 2 παρουσιάζονται εργασίες και μελέτες που έχουν προηγηθεί και είναι σχετικές με τα θέματα που καλύπτει η συγκεκριμένη πτυχιακή. Το Κεφάλαιο 3 αναλύει την διαδικασία επιλογής και εξόρυξης των δεδομένων και την διαδικασία επεξεργασίας τους. Τέλος, στο Κεφάλαιο 4 παρουσιάζεται μια σύνοψη της διαδικασίας που ακολουθήθηκε και στο Κεφάλαιο 5 παραθέτουμε τα συμπεράσματα από τη μελέτη αυτή καθώς και τις προτεινόμενες μελλοντικές επεκτάσεις.

2 *Σχετικές δραστηριότητες και εργασίες*

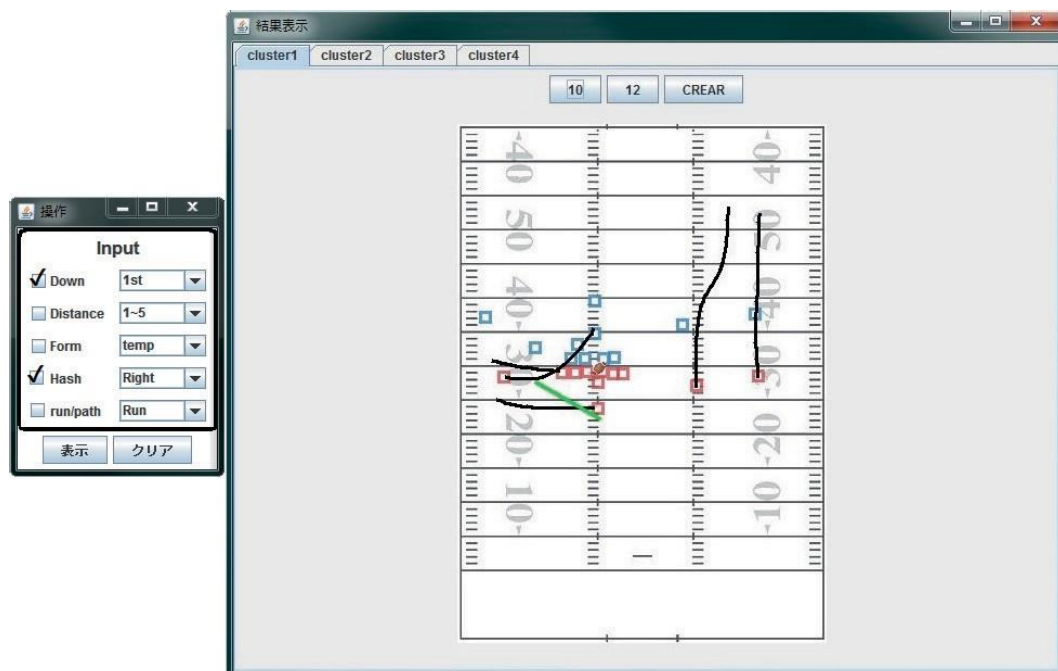
Τα δεδομένα και τα αναλυτικά στοιχεία αποτελούν μέρος της αθλητικής βιομηχανίας ήδη από τη δεκαετία του 1870, όταν καταγράφηκε το πρώτο boxscore στο μπέιζμπολ. Ωστόσο, μόλις πρόσφατα χρησιμοποιήθηκαν προηγμένες τεχνικές εξόρυξης δεδομένων και μηχανικής μάθησης για τη διευκόλυνση της λειτουργίας αθλητικών franchise. Ενώ μέρος του λόγου σχετίζεται με την ικανότητα συλλογής πιο λεπτομερών δεδομένων, ένας εξίσου σημαντικός παράγοντας για αυτήν τη στροφή προς τα αναλυτικά είναι η τεράστια επιτυχία και το ανταγωνιστικό πλεονέκτημα που απολάμβαναν οι πρώτοι υιοθέτες των επενδύσεων στα αναλυτικά στοιχεία (που διαδίδονται από τον best-seller » "Moneyball" που περιέγραψε την επιτυχία που είχε το Oakland Athletics με τα analytics).

2.1 *Οπτικοποίηση αθλητικών συστημάτων μέσω δεδομένων εξόρυξης τροχιάς*

Ένα σημαντικό καθήκον για τη βελτίωση της απόδοσης μιας αθλητικής ομάδας είναι η ανάλυση προηγούμενων αγώνων από διάφορες απόψεις. Υπάρχουν δύο τύποι μεθόδων ανάλυσης στρατηγικής αθλητισμού που αντιμετωπίζονται στα παιχνίδια με μπάλα: στατιστική ανάλυση και ανάλυση κίνησης. Στη στατιστική ανάλυση, η βαθμολόγηση πληροφοριών αναλύεται με στατιστικές μεθόδους, ενώ στην ανάλυση κίνησης, η κίνηση των παικτών και η μπάλα αναλύονται. Ωστόσο, δεν υπάρχει μια γενική μέθοδος και για τους δύο τύπους ανάλυσης. Επομένως, σε αυτήν την έρευνα, παρουσιάζεται ένα σύστημα υποστήριξης αποφάσεων για

Ανάλυση δεδομένων αγώνων μπάσκετ
Σπύρος Κανελλόπουλος

αθλητικές στρατηγικές για παιχνίδια με μπάλα στο οποίο και τα αποτελέσματα της στατιστικής ανάλυσης αλλά και των κινήσεων των παικτών απεικονίζονται με ολοκληρωμένο τρόπο για να διευκολύνεται η εύκολη λήψη αποφάσεων. Το σύστημά αυτό, έχει αναπτυχθεί για να υποστηρίξει στρατηγικές αποφάσεις που συνήθως λαμβάνονται από έναν προπονητή ομάδας. Η δομή αυτή, προτείνει επίσης και αναπτύσσει μια πιο αποτελεσματική μέθοδο εξόρυξης που επιτρέπει στον προπονητή να απεικονίζει δυναμικά την κίνηση.



Εικόνα 3: Οπτικοποίηση συστήματος ποδοσφαίρου μέσω διαγράμματος τροχιάς

2.2 Εξαγωγή και βαθμονόμηση θερμικών χαρτών από καταγραφές αθλητικών εφαρμογών

Η χρήση δεδομένων κίνησης από εφαρμογές παρακολούθησης αθλητικών κινήτων επηρεάζεται από τις εγγενείς προκαταλήψεις και ευαισθησία της, οι οποίες πρέπει να γίνουν κατανοητές κατά την ανάπτυξη υπηρεσιών προστιθέμενης αξίας, για παράδειγμα, χρηστών εφαρμογών και σχεδιαστών πόλεων. Έχουμε αναπτύξει μια μέθοδο για τη δημιουργία ενός χάρτη θερμότητας διατήρησης της ιδιωτικής

ζωής με την ποικιλομορφία των χρηστών (ppDIV), στην οποία λαμβάνεται υπόψη η πυκνότητα των τροχιών, καθώς και η ποικιλομορφία των χρηστών, αποτρέποντας έτσι τις επιπτώσεις προκατάληψης που προκαλούνται από την ανισότητα συμμετοχής. Η μέθοδος εφαρμόζεται σε δημόσιες προπονήσεις ποδηλασίας και συγκρίνεται με την εκτίμηση της πυκνότητας του πυρήνα που διατηρεί το απόρρητο (ppKDE) εστιάζοντας μόνο στην πυκνότητα των καταγεγραμμένων πορειών και στον υπολογισμό του αριθμού των χρηστών που διατηρούν το απόρρητο (ppUCC), που είναι παρόμοιος με τον αριθμό τετραγώνων του ατόμου χρήστες εφαρμογών. Μια συνειδητοποίηση της ιδιωτικής ζωής εισήχθη σε όλες τις μεθόδους ως βήμα προεπεξεργασίας δεδομένων ακολουθώντας την αρχή του k-Anonymity. Τα αποτελέσματα βαθμονόμησης για τους χάρτες θερμότητας που χρησιμοποιούν δεδομένα καταμέτρησης ποδηλάτων που συγκεντρώθηκαν από την πόλη του Ελσίνκι είναι καλά ($R^2 > 0,7$) και αυξάνουν τις υψηλές προσδοκίες για τη χρήση χαρτών θερμότητας σε ένα πλαίσιο πολεοδομικού σχεδιασμού. Αυτό υποστηρίζεται περαιτέρω από την ημερήσια κατανομή των προπονήσεων που δείχνει ότι, εκτός από τους αθλητικούς ποδηλάτες, πολλοί χρηστικοί ποδηλάτες παρακολουθούν τις μετακινήσεις τους. Ωστόσο, τα αθλητικά δεδομένα παρακολούθησης μπορούν να εμπλουτίσουν μόνο τις επίσημες επιτόπιες μετρήσεις με την υψηλή χωροχρονική ανάλυση και κάλυψη και όχι να τις αντικαταστήσουν.

2.3 Προσδιορισμός και κατανόηση αστικών αθλητικών περιοχών χρησιμοποιώντας το Nokia Sports Tracker

Οι τρέχουσες εξελίξεις στις διαδεδομένες τεχνολογίες επιτρέπουν στους χρήστες να δημιουργούν και να μοιράζονται έναν αυξανόμενο όγκο δεδομένων που βρίσκονται Έτσι, ορισμένα πλούσια σύνολα δεδομένων για την ανθρώπινη κινητικότητα καθίστανται διαθέσιμα στον Ιστό. Σε αυτό το άρθρο εξάγουμε περίπου 790.000 ίχνη κινητικότητας από ένα διαδικτυακό αποθετήριο κομματιών GPS - το Nokia Sports Tracker Service. Χρησιμοποιώντας μηχανισμούς εξόρυξης δεδομένων, δείχνουμε ότι αυτά τα δεδομένα μπορούν να αναλυθούν για να αποκαλύψουν καθημερινές ρουτίνες και ενδιαφέροντα σχήματα στη χρήση δημόσιων χώρων.

Ανάλυση δεδομένων αγώνων μπάσκετ
Σπύρος Κανελλόπουλος

Πρώτα δείχνουμε ότι η προσέγγισή μας υποστηρίζει την ευρείας κλίμακας ανάλυση του τόπου των ανθρώπων συγκρίνοντας τα πρότυπα συμπεριφοράς μεταξύ των πόλεων. Στη συνέχεια, χρησιμοποιώντας την εκτίμηση πυκνότητας πυρήνα, παρουσιάζουμε έναν μηχανισμό για τον εντοπισμό δημοφιλών αθλητικών περιοχών σε μεμονωμένες πόλεις. Αυτό το είδος ανάλυσης μάς επιτρέπει να αναδείξουμε ανθρωποκεντρικές γεωγραφίες που μπορούν να υποστηρίξουν ένα ευρύ φάσμα εφαρμογών, από υπηρεσίες βάσει τοποθεσίας έως πολεοδομικό σχεδιασμό.

2.4 Διερεύνηση πολυπλοκότητας της ανθρώπινης κίνησης διαμέσου δεδομένων τροχιάς

Οι τεχνολογίες των κινητών επικοινωνιών διαπερνούν την κοινωνία μας και τα ασύρματα δίκτυα αισθάνονται την κίνηση των ανθρώπων, δημιουργώντας μεγάλους όγκους δεδομένων κινητικότητας, όπως εγγραφές κλήσεων κινητής τηλεφωνίας και κομμάτια Global Positioning System (GPS). Σε αυτό το έργο, απεικονίζουμε την εντυπωσιακή αναλυτική δύναμη των μαζικών συλλογών δεδομένων τροχιάς στην αποκάλυψη της πολυπλοκότητας της ανθρώπινης κινητικότητας. Παρουσιάζουμε τα αποτελέσματα ενός πειράματος μεγάλης κλίμακας, με βάση τις λεπτομερείς διαδρομές δεκάδων χιλιάδων ιδιωτικών αυτοκινήτων με ενσωματωμένους δέκτες GPS, που παρακολουθούνται κατά τη διάρκεια εβδομάδων συνηθισμένης κινητής δραστηριότητας. Απεικονίζουμε τη διαδικασία ανακάλυψης γνώσης που, με βάση αυτά τα δεδομένα, εξετάζει ορισμένα θεμελιώδη ερωτήματα των αναλυτών κινητικότητας: ποια είναι τα συχνά πρότυπα των ταξιδιών των ανθρώπων; Πόσο μεγάλοι ελκυστήρες και εκπληκτικά γεγονότα επηρεάζουν την κινητικότητα; Πώς να προβλέψω περιοχές πυκνής κυκλοφορίας στο εγγύς μέλλον; Πώς να χαρακτηρίσετε την κυκλοφοριακή συμφόρηση και τη συμφόρηση; Περιγράφουμε επίσης το M-Atlas, τη γλώσσα και το σύστημα ερωτήσεων και εξόρυξης που καθιστά δυνατή αυτήν την αναλυτική διαδικασία, παρέχοντας τους μηχανισμούς για να κυριαρχήσουμε την πολυπλοκότητα της μετατροπής ακατέργαστων κομματιών GPS σε γνώσεις

**Ανάλυση δεδομένων αγώνων μπάσκετ
Σπύρος Κανελλόπουλος**

κινητικότητας. Το M-Atlas επικεντρώνεται στην ιδέα μιας τροχιάς και η διαδικασία ανακάλυψης γνώσης κινητικότητας μπορεί να προσδιοριστεί από ερωτήματα M-Atlas που πραγματοποιούν μετασχηματισμούς δεδομένων, εκτίμηση βάσει παραμέτρων των παραμέτρων των μεθόδων εξόρυξης, αξιολόγηση ποιότητας των ληφθέντων αποτελεσμάτων , η ποσοτική και οπτική εξερεύνηση των ανακαλυφθέντων μοντέλων συμπεριφοράς και μοντέλων, η σύνθεση των εξορυκτικών μοτίβων, μοντέλων και δεδομένων με περαιτέρω αναλύσεις και εξόρυξη, καθώς και οι στοιχειώδεις στρατηγικές εξόρυξης για την αντιμετώπιση της κλιμάκωσης.

3

Τεχνολογικό υπόβαθρο

Όπως έχει προαναφερθεί, ο σκοπός της παρούσας πτυχιακής εργασίας είναι η αποθήκευση αθλητικών στατιστικών στοιχείων, με στόχο την άντληση εξειδικευμένων πληροφοριών. Για την επίτευξη του παραπάνω στόχου, η μελέτη χωρίστηκε σε τρία στάδια, καθένα από τα οποία απαιτούσε λεπτομερή προσοχή σε κάθε βήμα με σκοπό τη βέλτιστη λειτουργία και απόδοση του αλγορίθμου τόσο σε επίπεδο χρόνου εκτέλεσης αλλά και ποιοτικών αποτελεσμάτων.

Αρχικά, με δεδομένο ότι είναι μια διαδικτυακή εφαρμογή, έπρεπε να στηθεί το κατάλληλο περιβάλλον που να στηρίζει την δομή αυτή. Έπειτα, έγινε μια συλλογή δεδομένων με στατιστικά ομάδων και παικτών του nba (National Basketball Federation). Στη συνέχεια, γίνεται ο διαχωρισμός με βάση την καταλληλότητά τους και η σύγκρισή τους με στόχο την εξόρυξη σύνθετων στατιστικών. Τέλος, γίνεται ανάλυση όλων των παραπάνω σταδίων με παράθεση όλων των ανωτέρω πληροφοριών και δίνοντας λεπτομερή έμφαση στην κατανόηση τους από τον αναγνώστη της παρούσας εργασίας.

3.1 Στάδιο δημιουργίας ψηφιακού περιβάλλοντος

Το ψηφιακό περιβάλλον είναι ένα περιβάλλον που χρησιμοποιείται από το Django για την εκτέλεση μιας εφαρμογής. Συνιστάται η δημιουργία και η εκτέλεση μιας εφαρμογής Django σε ξεχωριστό περιβάλλον. Η αιτία του διαχωρισμού αυτού έγκειται στην διαφάνεια και στην αποδοτικότητα της εφαρμογής. Ειδικότερα, η κάθε εφαρμογή που βρίσκεται σε ένα ξεχωριστό περιβάλλον χρησιμοποιεί τις

βιβλιοθήκες του, ανεξάρτητα από τις βιβλιοθήκες του τοπικού μηχανήματος. Η γλώσσα προγραμματισμού Python, παρέχει ένα εργαλείο `virtualenv` για τη δημιουργία ενός απομονωμένου περιβάλλοντος Python. Θα χρησιμοποιήσουμε αυτό το εργαλείο για να δημιουργήσουμε ένα εικονικό περιβάλλον για την εφαρμογή Django.

Για την δημιουργία ενός εικονικού περιβάλλοντος είναι απαραίτητη η χρήση ενός τερματικού λειτουργίας σε περιβάλλον linux, στο οποίο εκτελέστηκε η παρούσα έρευνα, ή γραμμής εντολών σε περιβάλλον windows. Ωστόσο, σε τερματικό linux είναι πιο εύκολη η λειτουργία και η συντήρηση μιας django εφαρμογής αφού πολλές βιβλιοθήκες της γλώσσας Python είναι περισσότερο συμβατές σε αυτό το περιβάλλον. Επιπρόσθετα, εξαιτίας της δομής του ψηφιακού περιβάλλοντος και του περιβάλλοντος linux είναι πιο γρήγορη και αποδοτική η λειτουργία του σε αυτό, με δεδομένο ότι καταναλώνει λιγότερους πόρους.

Αναλυτικότερα, στον home φάκελο πληκτρολογούμε την εντολή `$ apt-get install python3-venv` για να εγκαταστήσουμε τις απαραίτητες βιβλιοθήκες. Έπειτα, δημιουργούμε έναν φάκελο με την εντολή `$ mkdir djangoenv` (djangoenv είναι το όνομα του φακέλου, πληκτρολογούμε οποιοδήποτε χαρακτηριστικό όνομα επιθυμούμε) και εισερχόμαστε στον φάκελο αυτόν με την εντολή

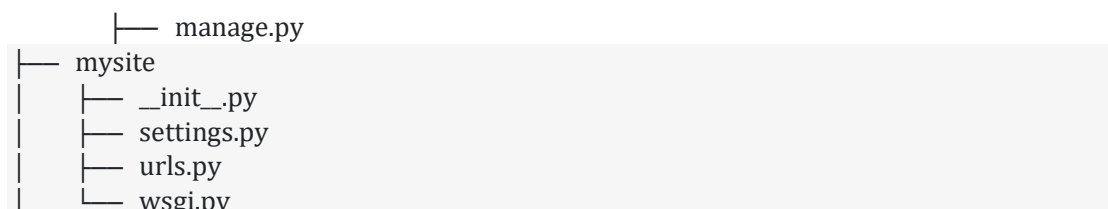
`$ cd djangoenv`. Για την δημιουργία του ψηφιακού περιβάλλοντος, πληκτρολογούμε `$ python3 -m venv djangoenv`. Επιπλέον, για την ενεργοποίηση του περιβάλλοντος, εισερχόμαστε στον υποφάκελο `bin` του φακέλου `djangoenv` και πληκτρολογούμε `$ source activate`. Τέλος, για την εγκατάσταση της django βιβλιοθήκης πληκτρολογούμε `$ pip install django`

3.2 Στάδιο οργάνωσης αρχείων εφαρμογής

Για την δημιουργία της εφαρμογής πρέπει να βρισκόμαστε στον φάκελο όπου είναι ενεργοποιημένο το ψηφιακό περιβάλλον και να πληκτρολογήσουμε την εντολή `$ django-admin startproject mysite`, όπου `mysite` είναι το όνομα του project. Με την εντολή αυτή δημιουργείται ένας φάκελος με το όνομα `mysite`, ο οποίος περιέχει την βασική δομή ενός django project. Η δομή αυτή, αποτελείται από ένα

σύνολο αρχείων, `manage.py`, `asgi.py`, `__init.py__`, `settings.py`, `urls.py`, `wsgi.py`, τα οποία έχουν το κάθε ένα την δική τους λειτουργία στο περιβάλλον django.

- `manage.py`: Περιέχει συγκεκριμένες παραμέτρους για την εκτέλεση της εφαρμογής
- `asgi.py`: Περιέχει ASGI(Asynchronous Server Gateway Interface) παραμέτρους για την επικοινωνία του web server με την django εφαρμογή.
- `__init.py__`: Γενικό Python αρχείο που αναφέρει τα ονόματα των βιβλιοθηκών που έχουν συμπεριληφθεί (mysql database)
- `settings.py`: Αρχείο που περιέχει τις παραμέτρους εκτέλεσης του django προγράμματος.
- `urls.py`: Περιέχει τα url links που χρησιμοποιούνται για να καλέσουμε τις αντίστοιχες προγραμματιστικές μεθόδους. Θα αναφερθούμε σε αυτό αργότερα.
- `wsgi.py`: Αρχείο που περιέχει wsgi (Web Server Gateway Interface) παραμέτρους για την επικοινωνία του web server με την εφαρμογή. Ο wsgi τρόπος επικοινωνίας αποτελεί πρόγονο του asgi τρόπου επικοινωνίας



Σχήμα 1: Σχεδιάγραμμα για την καλύτερη απεικόνιση της django δομής

Το επόμενο βήμα είναι η δημιουργία ενός φακέλου που θα περιέχει τον κώδικα της εφαρμογής. Για την δημιουργία του φακέλου χρησιμοποιούμε την εντολή `$ python manage.py startapp mysite`. Με την εκτέλεση της εντολής αυτής, παρατηρούμε την δημιουργία ενός φακέλου με το όνομα `mysite` που περιέχει τα εξής αρχεία `__init.py__`, `models.py`, `tests.py`, `views.py`. Τα επιπρόσθετα αρχεία που δημιουργήθηκαν έχουν τις εξής λειτουργίες.

- `models.py`: Περιέχει τα αντικείμενα (`models`), τα πεδία τους και συγκεκριμένες μεθόδους που αφορούν τα αντικείμενα αυτά.
- `tests.py`: Περιέχει κάποια testing σενάρια με στόχο την πρόληψη λαθών κώδικα.
- `views.py`: Αποτελεί ένα από τα πιο σημαντικά αρχεία. Περιέχει τις μεθόδους οι οποίες χρησιμοποιούν τα αντικείμενα του `models.py` και τα επεξεργάζονται.

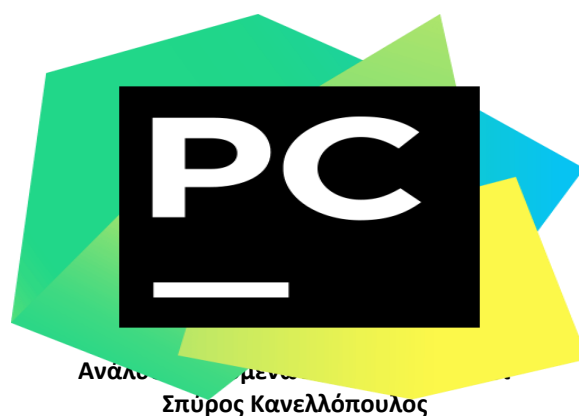
```

├── manage.py
├── mysite
│   ├── __init__.py
│   ├── settings.py
│   ├── urls.py
│   └── wsgi.py
├── mysite
│   ├── __init__.py
│   ├── models.py
│   ├── tests.py
│   └── views.py

```

Σχήμα 2: Τελικό σχεδιάγραμμα *django* δομής

Έπειτα, για την υλοποίηση κώδικα χρησιμοποιούμε το γραφικό περιβάλλον του Pycharm IDE που εξειδικεύεται στην γλώσσα προγραμματισμού Python, της εταιρείας JetBrains.



3.3 Στάδιο συλλογής αθλητικών στατιστικών δεδομένων

Για το στάδιο της συλλογής των στατιστικών δεδομένων χρησιμοποιήθηκαν πληθώρα ιστοσελίδων ωστόσο βρήκαμε την πιο αξιόπιστη και ολοκληρωμένη πηγή στο Basketball Reference (<https://www.basketball-reference.com>). Οι πληροφορίες που συλλέχθηκαν αφορούν τα συνολικά στατιστικά των ομάδων αλλά και τα ατομικά στατιστικά των παικτών. Στόχος αυτών, είναι η εξαγωγή σύνθετων συμπερασμάτων, τα οποία δίνουν την δυνατότητα στο ανθρώπινο δυναμικό που τα διαχειρίζεται να λάβει αποφάσεις ευκολότερα. Ειδικότερα, ορισμένα δεδομένα αφορούσαν τα προσωπικά στοιχεία των παικτών όπως είναι το ύψος, βάρος, ημερομηνία γέννησης. Οι πληροφορίες αυτές χρησιμοποιήθηκαν για την κατηγοριοποίηση τους.

Επιπλέον τα κυριότερα δεδομένα αποτελούν τα ποσοστά που έχουν οι παίκτες σε καλάθια τριών (τρίποντα) και δύο (δίποντα) πόντων. Απαραίτητα χαρακτηριστικά αποτελούν επίσης οι επιθετικές και αμυντικές διεκδικήσεις (offensive & defensive rebounds), για την ανάδειξη της ενεργητικότητας της ομάδας σε άμυνα και επίθεση αντίστοιχα. Κυρίαρχο στοιχείο αποτελούν οι τελικές πάσες (assists) που δίνονται ανά αγώνα. Τέλος, ενδεικτικό χαρακτηριστικό αποτελούν τα λάθη (turnovers) που κάνει μια ομάδα.

Στην παρούσα έρευνα, τα ανωτέρω χαρακτηριστικά χρησιμοποιήθηκαν στο αρχείο models.py. Ειδικότερα, δημιουργήθηκαν οι παρακάτω κλάσεις τύπου model, με στόχο την εύκολη και γρήγορη αποθήκευση πολλαπλών αντικειμένων.

- Teams: κλάση που αποθηκεύονται οι ομάδες
 - 1 team_name: Μεταβλητή τύπου character στην οποία βάζουμε το όνομα της ομάδας που επιθυμούμε

Χαρακτηριστικά	Περιγραφή
Team name	Το όνομα της ομάδας

Πίνακας 1: Χαρακτηριστικά Κλάσης Team

- Players: κλάση που αποθηκεύονται οι παίκτες
 - 1 number: Μεταβλητή τύπου integer στην οποία βάζουμε το νούμερο που θα έχει ο αθλητής στην φανέλα
 - 2 firstname: Μεταβλητή τύπου character στην οποία εισάγουμε το μικρό όνομα του αθλητή
 - 3 lastname: Μεταβλητή τύπου character στην οποία εισάγουμε το επώνυμο του αθλητή
 - 4 dob dd/mm/year(date of birth): Μεταβλητή τύπου date στην οποία εισάγουμε την ημερομηνία γέννησης του αθλητή
 - 5 height: Μεταβλητή τύπου character στην οποία τοποθετούμε το ύψος του αθλητή σε πόδια.
 - 6 weight: Μεταβλητή τύπου integer στην οποία τοποθετούμε το βάρος του αθλητή σε πόδια.
 - 7 position: Μεταβλητή τύπου character τα οποία ωστόσο έχουν προκαθορισμένες τιμές. Στις τιμές αυτές επιλέγουμε ποιά θέση έχει ο αθλητής στο γήπεδο
 - 8 dominant_hand: Μεταβλητή τύπου character τα οποία ωστόσο έχουν προκαθορισμένες τιμές. Στις τιμές αυτές επιλέγουμε ποιά είναι το κυρίαρχο χέρι του αθλητή.
 - 9 nationality: Μεταβλητή τύπου character στην οποία εισάγουμε την εθνικότητα του παίκτη
 - 10 experience: Μεταβλητή τύπου integer στην οποία εισάγουμε τα χρόνια εμπειρίας που έχει ο αθλητής στον συγκεκριμένο φορέα (nba).

11 teamID: Μεταβλητή τύπου ξένου κλειδιού που αντιστοιχεί στον πίνακα που περιέχονται οι ομάδες. Σε αυτήν την μεταβλητή καταχωρούμε σε ποιά ομάδα ανήκει ο αθλητής.

**Στα πεδία teamID, dominant_hand και position ορίζουμε εμείς το εύρος των τιμών με ένα dropdown με διαθέσιμες επιλογές.

Χαρακτηριστικά	Περιγραφή
Number	Ο αριθμός που φορά ο αθλητής στην φανέλα
Firstname	Το μικρό όνομα του αθλητή
Lastname	Το επώνυμο του αθλητή
Dob	Η ημερομηνία γέννησης του αθλητή
Height	Το ύψος σε πόδια (ft)
Weight	Το βάρος σε λίβρες (lb)
Position	Η θέση που έχει ο αθλητής στο γήπεδο, διακρίνεται playmaker, point guard, small forward, power forward, center
Dominant Hand	Κυρίαρχο χέρι, δεξιόχειρας ή αριστερόχειρας
Nationality	Ιθαγένεια
Experience	Χρόνια εμπειρίας στο nba
TeamID	Η ομάδα στην οποία ανήκει

Πίνακας 2: Χαρακτηριστικά Κλάσης Player

- TeamStats: Κλάση στην οποία αποθηκεύονται τα ομαδικά στατιστικά
 - 1 games: Μεταβλητή τύπου Integer που δηλώνει πόσα παιχνίδια έχει παίξει η ομάδα
 - 2 field_goals: Μεταβλητή τύπου Integer που δηλώνει πόσα καλάθια έχει βάλει μια ομάδα σε μια χρονιά.
 - 3 field_goal_attempts: Μεταβλητή τύπου Integer που δηλώνει πόσες προσπάθειες για καλάθια έχει επιχειρήσει να βάλει μια ομάδα σε μια χρονιά.
 - 4 field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει το ποσοστό ευστοχίας μια ομάδας σε μια χρονιά.
 - 5 three_point_field_goal: Μεταβλητή τύπου Integer που δηλώνει πόσα τρίποντα καλάθια έχει βάλει μια ομάδα σε μια χρονιά.
 - 6 three_point_field_goal_attempts: Μεταβλητή τύπου Integer που δηλώνει πόσες τρίποντες προσπάθειες για καλάθια έχει επιχειρήσει να βάλει μια ομάδα σε μια χρονιά.
 - 7 three_point_field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει το ποσοστό ευστοχίας της ομάδας στα τρίποντα καλάθια.
 - 8 two_point_field_goal: Μεταβλητή τύπου Integer που δηλώνει πόσα δίποντα καλάθια έχει βάλει μια ομάδα σε μια χρονιά.
 - 9 two_point_field_goal_attempts: Μεταβλητή τύπου Integer που δηλώνει πόσες δίποντες προσπάθειες για καλάθια έχει επιχειρήσει να βάλει μια ομάδα σε μια χρονιά.
 - 10 two_point_field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει το ποσοστό ευστοχίας της ομάδας στα δίποντα καλάθια.
 - 11 free_throws: Μεταβλητή τύπου Integer που δηλώνει πόσες ελεύθερες βολές έχει βάλει μια ομάδα σε μια χρονιά.
 - 12 free_throw_attempts: Μεταβλητή τύπου Integer που δηλώνει πόσες προσπάθειες σε ελεύθερες βολές έχει επιχειρήσει να βάλει μια ομάδα σε μια χρονιά.

- 13 free_throw_percentage: Μεταβλητή τύπου Decimal που δηλώνει το ποσοστό ευστοχίας της ομάδας στα δίποντα καλάθια.
- 14 offensive_rebounds: Μεταβλητή τύπου Integer που δηλώνει τις επιθετικές διεκδικήσεις μιας ομάδας.
- 15 defensive_rebounds: Μεταβλητή τύπου Integer που δηλώνει τις αμυντικές διεκδικήσεις μιας ομάδας.
- 16 total_rebounds: Μεταβλητή τύπου Integer που δηλώνει τις συνολικές διεκδικήσεις μιας ομάδας.
- 17 assists: Μεταβλητή τύπου Integer που δηλώνει τις τελικές πάσες μιας ομάδας.
- 18 steals: Μεταβλητή τύπου Integer που δηλώνει τα κλεψίματα μιας ομάδας.
- 19 turnovers: Μεταβλητή τύπου Integer που δηλώνει τα λάθη μιας ομάδας.
- 20 blocks: Μεταβλητή τύπου Integer που δηλώνει τα κοψίματα μιας ομάδας.
- 21 personal_fouls: Μεταβλητή τύπου Integer που δηλώνει τα foul μιας ομάδας.
- 22 points: Μεταβλητή τύπου Integer που δηλώνει τους συνολικούς πόντους μιας ομάδας.
- 23 teamID: Μεταβλητή τύπου Foreign Key που δηλώνει σε ποιά ομάδα ανήκουν τα στατιστικά αυτά.

Χαρακτηριστικά	Περιγραφή
Games	Ο αριθμός αγώνων του αθλητή
Field Goals	Τα συνολικά καλάθια που έχει βάλει
Field Goal Attempts	Ο αριθμός των προσπαθειών για καλάθι

Field Goal Percentage	Το ποσοστό που προέκυψε από Field goals / field goal attempts
Three Point Field Goals	Τα τρίποντα καλάθια
Three Point Field Goal Attempts	Οι τρίποντες προσπάθειες του αθλητή
Three Point Field Goal Percentage	Το ποσοστό που προέκυψε από three point field goals / field goal attempts
Two Point Field Goals	Τα δίποντα καλάθια
Two Point Field Goal Attempts	Οι δίποντες προσπάθειες του αθλητή
Two Point Field Goal Percentage	Το ποσοστό που προέκυψε από two point field goals / field goal attempts
Free Throws	Οι εύστοχες ελεύθερες βολές
Free Throw Attempts	Οι συνολικές ελεύθερες βολές
Free Throw Percentage	Ποσοστό ευστοχίας ελεύθερων βολών
Offensive Rebounds	Επιθετικές Διεκδικήσεις
Defensive Rebounds	Αμυντικές Διεκδικήσεις
Total Rebounds	Συνολικές Διεκδικήσεις
Assists	Τελικές πάσες
Steals	Κλεψίματα
Turnovers	Λάθη

Blocks	Κοψίματα
Personal Fouls	Προσωπικά Φάουλ
Points	Πόντοι
TeamID	Το id της ομάδας για να γίνει η στοίχιση

Πίνακας 3: Χαρακτηριστικά Κλάσης TeamStats

- PlayerStats: Κλάση στην οποία αποθηκεύονται τα ατομικά στατιστικά
 - 1.1 games: Μεταβλητή τύπου Integer που δηλώνει πόσα παιχνίδια έχει παίξει ο αθλητής
 - 1.2 minutes_played: Μεταβλητή τύπου Decimal που δηλώνει πόσα λεπτά συμμετοχής έχει ανά αγώνα
 - 1.3 field_goals: Μεταβλητή τύπου Decimal που δηλώνει πόσα καλάθια βάζει ο αθλητής ανά αγώνα
 - 1.4 field_goal_attempts: Μεταβλητή τύπου Decimal που δηλώνει πόσες προσπάθειες για καλάθια έχει επιχειρεί να βάλει ένας παίκτης ανά αγώνα.
 - 1.5 field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει το ποσοστό ευστοχίας ενός αθλητή ανά αγώνα.
 - 1.6 three_point_field_goal: Μεταβλητή τύπου Decimal που δηλώνει πόσα τρίποντα καλάθια βάζει ένας αθλητής ανά αγώνα
 - 1.7 three_point_field_goal_attempts: Μεταβλητή τύπου Decimal που δηλώνει πόσες τρίποντες προσπάθειες για καλάθια επιχειρεί να βάλει ένας αθλητής ανά αγώνα.
 - 1.8 three_point_field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει το ποσοστό ευστοχίας του αθλητή στα τρίποντα καλάθια ανά αγώνα.

- 1.9 two_point_field_goal: Μεταβλητή τύπου Decimal που δηλώνει πόσα δίποντα καλάθια βάζει ένας αθλητής.
- 1.10 two_point_field_goal_attempts: Μεταβλητή τύπου Decimal που δηλώνει πόσες δίποντες προσπάθειες για καλάθια επιχειρεί.
- 1.11 two_point_field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει το ποσοστό ευστοχίας του αθλητή στα δίποντα καλάθια.
- 1.12 free_throws: Μεταβλητή τύπου Decimal που δηλώνει πόσες ελεύθερες βολές βάζει.
- 1.13 free_throw_attempts: Μεταβλητή τύπου Decimal που δηλώνει πόσες προσπάθειες σε ελεύθερες βολές επιχειρεί να βάλει.
- 1.14 free_throw_percentage: Μεταβλητή τύπου Decimal που δηλώνει το ποσοστό ευστοχίας του αθλητή στα δίποντα καλάθια.
- 1.15 offensive_rebounds: Μεταβλητή τύπου Decimal που δηλώνει τις επιθετικές διεκδικήσεις του αθλητή.
- 1.16 defensive_rebounds: Μεταβλητή τύπου Decimal που δηλώνει τις αμυντικές διεκδικήσεις του αθλητή.
- 1.17 total_rebounds: Μεταβλητή τύπου Decimal που δηλώνει τις συνολικές διεκδικήσεις του αθλητή.
- 1.18 assists: Μεταβλητή τύπου Decimal που δηλώνει τις τελικές πάσες μιας ομάδας.
- 1.19 steals: Μεταβλητή τύπου Decimal που δηλώνει τα κλεψίματα.
- 1.20 turnovers: Μεταβλητή τύπου Decimal που δηλώνει τα λάθη του αθλητή.
- 1.21 blocks: Μεταβλητή τύπου Decimal που δηλώνει τα κοψίματα.
- 1.22 personal_fouls: Μεταβλητή τύπου Decimal που δηλώνει τα foul.
- 1.23 points: Μεταβλητή τύπου Decimal που δηλώνει τους συνολικούς πόντους.
- 1.24 playerId: Μεταβλητή τύπου Foreign Key που δηλώνει σε ποιόν αθλητή αντιστοιχούν τα στατιστικά αυτά.

**** Οι μεταβλητές της κλάσης PlayerStats αντιστοιχούν σε κάθε αγώνα, λχ πόσους πόντους βάζει ένας παίκτης ανά αγώνα.**

Χαρακτηριστικά	Περιγραφή
Games	Ο αριθμός αγώνων του αθλητή
Field Goals	Τα συνολικά καλάθια που έχει βάλει
Field Goal Attempts	Ο αριθμός των προσπαθειών για καλάθι
Field Goal Percentage	Το ποσοστό που προέκυψε από Field goals / field goal attempts
Three Point Field Goals	Τα τρίποντα καλάθια
Three Point Field Goal Attempts	Οι τρίποντες προσπάθειες του αθλητή
Three Point Field Goal Percentage	Το ποσοστό που προέκυψε από three point field goals / field goal attempts
Two Point Field Goals	Τα δίποντα καλάθια
Two Point Field Goal Attempts	Οι δίποντες προσπάθειες του αθλητή
Two Point Field Goal Percentage	Το ποσοστό που προέκυψε από two point field goals / field goal attempts
Free Throws	Οι εύστοχες ελεύθερες βολές
Free Throw Attempts	Οι συνολικές ελεύθερες βολές
Free Throw Percentage	Ποσοστό ευστοχίας ελεύθερων βολών
Offensive Rebounds	Επιθετικές Διεκδικήσεις

Defensive Rebounds	Αμυντικές Διεκδικήσεις
Total Rebounds	Συνολικές Διεκδικήσεις
Assists	Τελικές πάσες
Steals	Κλεψίματα
Turnovers	Λάθη
Blocks	Κοψίματα
Personal Fouls	Προσωπικά Φάουλ
Points	Πόντοι
PlayerID	Το id του αθλητή για να γίνει η στοίχιση

Πίνακας 4: Χαρακτηριστικά Κλάσης *PlayerStats*


```

79 class PlayerStats(models.Model):
80     games = models.IntegerField(validators=[MaxValueValidator(110)], default=0)
81     minutes_played = models.DecimalField(max_digits=5, decimal_places=2, null=True)
82
83     # type is decimal cause it's per game stats
84     field_goals = models.DecimalField(max_digits=5, decimal_places=2, null=True)
85     field_goal_attempts = models.DecimalField(max_digits=5, decimal_places=2, null=True)
86     field_goal_percentage = models.DecimalField(max_digits=5, decimal_places=2, null=True)
87
88     three_point_field_goals = models.DecimalField(max_digits=5, decimal_places=2, null=True)
89     three_point_field_goal_attempts = models.DecimalField(max_digits=5, decimal_places=2, null=True)
90     three_point_field_goal_percentage = models.DecimalField(max_digits=5, decimal_places=2, null=True)
91
92     two_point_field_goals = models.DecimalField(max_digits=5, decimal_places=2, null=True)

```

Απόσπασμα κώδικα 1: Ορισμός κλάσης *PlayerStats*

- **AverageTeamStats:** Κλάση στην οποία αποθηκεύονται ο μέσος όρος κάθε πεδίου των ομάδων
 - 1 **games:** Μεταβλητή τύπου *Decimal* που δηλώνει τον μέσο όρο των παιχνιδιών μιας ομάδας.
 - 2 **field_goals:** Μεταβλητή τύπου *Decimal* που δηλώνει τον μέσο όρο καλαθιών έχει βάλει μιας ομάδας σε μια χρονιά.
 - 3 **field_goal_attempts:** Μεταβλητή τύπου *Decimal* που δηλώνει τον μέσο όρο των προσπαθειών για καλάθια που έχει επιχειρήσει να βάλει μια ομάδα σε μια χρονιά.
 - 4 **field_goal_percentage:** Μεταβλητή τύπου *Decimal* που δηλώνει τον μέσο όρο του ποσοστού ευστοχίας μιας ομάδας σε μια χρονιά.

- 5 three_point_field_goal: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των τρίποντων καλαθιών που έχει βάλει μια ομάδα σε μια χρονιά.
- 6 three_point_field_goal_attempts: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των τρίποντων προσπαθειών για καλάθια που έχει επιχειρήσει να βάλει μια ομάδα σε μια χρονιά.
- 7 three_point_field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των ποσοστών ευστοχίας των ομάδων στα τρίποντα καλάθια.
- 8 two_point_field_goal: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των δίποντων καλαθιών που έχει βάλει μια ομάδα σε μια χρονιά.
- 9 two_point_field_goal_attempts: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των δίποντων προσπαθειών μιας ομάδας σε μια χρονιά.
- 10 two_point_field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο του ποσοστού ευστοχίας της ομάδας στα δίποντα καλάθια.
- 11 free_throws: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο εύστοχων ελεύθερων βολών μιας ομάδας σε μια χρονιά.
- 12 free_throw_attempts: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο προσπαθειών σε ελεύθερες βολές που επιχειρεί να βάλει μια ομάδα σε μια χρονιά.
- 13 free_throw_percentage: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο ποσοστού ευστοχίας μιας ομάδας στα δίποντα καλάθια.
- 14 offensive_rebounds: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο επιθετικών διεκδικήσεων μιας ομάδας.
- 15 defensive_rebounds: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο αμυντικών διεκδικήσεων μιας ομάδας.

- 16 total_rebounds: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο συνολικών διεκδικήσεων μιας ομάδας.
- 17 assists: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο τελικών πασών μιας ομάδας.
- 18 steals: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο κλεψιμάτων μιας ομάδας.
- 19 turnovers: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο λαθών μιας ομάδας.
- 20 blocks: Μεταβλητή τύπου Decimal που δηλώνει τα κοψίματα μιας ομάδας.
- 21 personal_fouls: Μεταβλητή τύπου Decimal που δηλώνει τα foul μιας ομάδας.
- 22 points: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο συνολικών πόντων μιας ομάδας.

Χαρακτηριστικά	Περιγραφή
Games	Ο αριθμός αγώνων του αθλητή
Field Goals	Τα συνολικά καλάθια που έχει βάλει
Field Goal Attempts	Ο αριθμός των προσπαθειών για καλάθι
Field Goal Percentage	Το ποσοστό που προέκυψε από Field goals / field goal attempts
Three Point Field Goals	Τα τρίποντα καλάθια
Three Point Field Goal Attempts	Οι τρίποντες προσπάθειες του αθλητή
Three Point Field Goal Percentage	Το ποσοστό που προέκυψε από three point field goals / field goal attempts

Two Point Field Goals	Τα δίποντα καλάθια
Two Point Field Goal Attempts	Οι δίποντες προσπάθειες του αθλητή
Two Point Field Goal Percentage	Το ποσοστό που προέκυψε από two point field goals / field goal attempts
Free Throws	Οι εύστοχες ελεύθερες βολές
Free Throw Attempts	Οι συνολικές ελεύθερες βολές
Free Throw Percentage	Ποσοστό ευστοχίας ελεύθερων βολών
Offensive Rebounds	Επιθετικές Διεκδικήσεις
Defensive Rebounds	Αμυντικές Διεκδικήσεις
Total Rebounds	Συνολικές Διεκδικήσεις
Assists	Τελικές πάσες
Steals	Κλεψίματα
Turnovers	Λάθη
Blocks	Κοψίματα
Personal Fouls	Προσωπικά Φάουλ
Points	Πόντοι

Πίνακας 5: Χαρακτηριστικά Κλάσης *AverageTeamStats*

- AveragePlayerStats: Κλάση στην οποία αποθηκεύονται ο μέσος όρος κάθε πεδίου των παικτών
 - 1 games: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των παιχνιδιών ενός παίκτη.
 - 2 field_goals: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο καλαθιών που έχει βάλει ένας παίκτης.
 - 3 field_goal_attempts: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των προσπαθειών για καλάθια που έχει επιχειρήσει να βάλει ένας παίκτης.
 - 4 field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο του ποσοστού ευστοχίας ένας παίκτης.
 - 5 three_point_field_goal: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των τρίποντων καλαθιών που έχει βάλει ένας παίκτης.
 - 6 three_point_field_goal_attempts: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των τρίποντων προσπαθειών για καλάθια που έχει επιχειρήσει να βάλει ένας παίκτης.
 - 7 three_point_field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των ποσοστών ευστοχίας ενός παίκτη στα τρίποντα καλάθια.
 - 8 two_point_field_goal: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των δίποντων καλαθιών που έχει βάλει ένας παίκτης.
 - 9 two_point_field_goal_attempts: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο των δίποντων προσπαθειών ένας παίκτης
 - 10 two_point_field_goal_percentage: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο του ποσοστού ευστοχίας ενός παίκτη στα δίποντα καλάθια.
 - 11 free_throws: Μεταβλητή τύπου Decimal που δηλώνει τον μέσο όρο εύστοχων ελεύθερων βολών ένας παίκτης.

- 12 `free_throw_attempts`: Μεταβλητή τύπου `Decimal` που δηλώνει τον μέσο όρο προσπαθειών σε ελεύθερες βολές που επιχειρεί να βάλει ένας παίκτης.
- 13 `free_throw_percentage`: Μεταβλητή τύπου `Decimal` που δηλώνει τον μέσο όρο ποσοστού ευστοχίας ενός παίκτη στα δίποντα καλάθια.
- 14 `offensive_rebounds`: Μεταβλητή τύπου `Decimal` που δηλώνει τον μέσο όρο επιθετικών διεκδικήσεων ενός παίκτη.
- 15 `defensive_rebounds`: Μεταβλητή τύπου `Decimal` που δηλώνει τον μέσο όρο αμυντικών διεκδικήσεων ενός παίκτη.
- 16 `total_rebounds`: Μεταβλητή τύπου `Decimal` που δηλώνει τον μέσο όρο συνολικών διεκδικήσεων ενός παίκτη.
- 17 `assists`: Μεταβλητή τύπου `Decimal` που δηλώνει τον μέσο όρο τελικών πασών ενός παίκτη.
- 18 `steals`: Μεταβλητή τύπου `Decimal` που δηλώνει τον μέσο όρο κλεψιμάτων ενός παίκτη.
- 19 `turnovers`: Μεταβλητή τύπου `Decimal` που δηλώνει τον μέσο όρο λαθών ενός παίκτη.
- 20 `blocks`: Μεταβλητή τύπου `Decimal` που δηλώνει τα κοψίματα ενός παίκτη.
- 21 `personal_fouls`: Μεταβλητή τύπου `Decimal` που δηλώνει τα foul ενός παίκτη.
- 22 `points`: Μεταβλητή τύπου `Decimal` που δηλώνει τον μέσο όρο συνολικών πόντων ενός παίκτη.

Χαρακτηριστικά	Περιγραφή
Games	Ο αριθμός αγώνων του αθλητή
Field Goals	Τα συνολικά καλάθια που έχει βάλει

Field Goal Attempts	Ο αριθμός των προσπαθειών για καλάθι
Field Goal Percentage	Το ποσοστό που προέκυψε από Field goals / field goal attempts
Three Point Field Goals	Τα τρίποντα καλάθια
Three Point Field Goal Attempts	Οι τρίποντες προσπάθειες του αθλητή
Three Point Field Goal Percentage	Το ποσοστό που προέκυψε από three point field goals / field goal attempts
Two Point Field Goals	Τα δίποντα καλάθια
Two Point Field Goal Attempts	Οι δίποντες προσπάθειες του αθλητή
Two Point Field Goal Percentage	Το ποσοστό που προέκυψε από two point field goals / field goal attempts
Free Throws	Οι εύστοχες ελεύθερες βολές
Free Throw Attempts	Οι συνολικές ελεύθερες βολές
Free Throw Percentage	Ποσοστό ευστοχίας ελεύθερων βολών
Offensive Rebounds	Επιθετικές Διεκδικήσεις
Defensive Rebounds	Αμυντικές Διεκδικήσεις
Total Rebounds	Συνολικές Διεκδικήσεις
Assists	Τελικές πάσες
Steals	Κλεψίματα
Turnovers	Λάθη

Blocks	Κοψίματα
Personal Fouls	Προσωπικά Φάουλ
Points	Πόντοι

Πίνακας 6: Χαρακτηριστικά Κλάσης AveragePlayerStats

3.4 Στάδιο αποθήκευσης αθλητικών στατιστικών δεδομένων

Αφού αναφερθήκαμε στον τρόπο με τον οποίο συλλέχθηκαν και οργανώθηκαν τα δεδομένα, πρέπει να αναφερθούμε και στον τρόπο στον οποίο αποθηκεύτηκαν τα δεδομένα της έρευνας αυτής. Οι πληροφορίες αυτές αποθηκεύτηκαν τοπικά με την χρήση μιας MySQL βάσης δεδομένων και για δική μας διευκόλυνσης ενός MySQL workbench για να χειριζόμαστε τα δεδομένα πιο γρήγορα. Για να δημιουργήσουμε την τοπική βάση δεδομένων πληκτρολογήσαμε την ακόλουθη εντολή στο τερματικό μας.

\$ pip install mysql. Η εντολή αυτή, εγκαθιστά της κατάλληλες βιβλιοθήκες για να τις χρησιμοποιήσουμε αργότερα. Ωστόσο δεν θα γίνει περισσότερη αναφορά στην τοπική βάση δεδομένων καθώς ξεφεύγει τους σκοπούς της έρευνας αυτής. Απεναντίας, θα αναφερθούμε στον τρόπο με τον οποίο συνδέθηκε η βάση δεδομένων με την εφαρμογή μας.

Για να επιτευχθεί σύνδεση μεταξύ της εφαρμογής και της τοπικής βάσης δεδομένων πρέπει να τοποθετήσουμε τα απαραίτητα διαπιστευτήρια στοιχεία της βάσης μας. Η τοποθεσία που πρέπει να υπάρχουν τα στοιχεία αυτά βρίσκεται στο αρχείο settings.py της web εφαρμογής μας. Ο λόγος που αποθηκεύονται εκεί, είναι διότι στο αρχείο αυτό αποθηκεύονται όλες οι απαραίτητες ρυθμίσεις που χρειάζεται μια διαδικτυακή εφαρμογή για να λειτουργήσει ομαλά. Τα διαπιστευτήρια της βάσης αναγράφονται με την μορφή json string και περιέχουν τα ακόλουθα στοιχεία:

- Engine: Αποτελεί τον τύπο βάσης δεδομένων, στην έρευνα χρησιμοποιήθηκε MySQL βάση.
- Name: Αποτελεί το σχήμα (schema) στο οποίο θα αποθηκευτούν τα δεδομένα.
- Host: Αποτελεί το url στο οποίο θα αποθηκευτούν τα δεδομένα μας, στην συγκεκριμένη έρευνα χρησιμοποιούμε localhost αφού αποθηκεύονται τοπικά.
- User: Αποτελεί το username του διαχειριστή της βάσης δεδομένων
- Password: Αποτελεί το συνθηματικό το οποίο πρέπει να χρησιμοποιήσει σε συνδυασμό με το username για να ταυτοποιηθεί ο χρήστης σαν διαχειριστής.
- Port: Αποτελεί της πύλη της ιστοσελίδας στην οποία θα σταλούν τα δεδομένα.

Απόσπασμα
μα
κώδικα 2:
Credential

```

1 DATABASES = {
2     'default': {
3         'ENGINE': 'django.db.backends.mysql',
4         'NAME': 'basket_stats',
5         'HOST': 'localhost',
6         'USER': 'root',
7         'PASSWORD': 'root_password',
8         'PORT': '3306'
9     }
10 }

```

ς βάσης δεδομένων

Αφού τοποθετήσαμε τις ρυθμίσεις για την βάση δεδομένων, χρειάζεται να τρέξουμε μια σειρά εντολών οι οποίες μας επιτρέπουν να συνδεθεί η βάση με την εφαρμογή. Οι εντολές αυτές είναι οι ακόλουθες:

1 \$ python manage.py makemigrations

Ανάλυση δεδομένων αγώνων μπάσκετ
Σπύρος Κανελλόπουλος

2 \$ python manage.py migrate

Για να λειτουργήσουν επιτυχημένα οι εντολές αυτές πρέπει να έχουμε ενεργοποιημένο το virtual environment και να βρισκόμαστε στον φάκελο όπου βρίσκεται το αρχείο manage.py. Κατά την εκτέλεση της πρώτης εντολής, αν είναι επιτυχημένη θα εμφανίσει ένα κατάλληλο μήνυμα και θα δημιουργήσει ένα ξεχωριστό αρχείο κάτω από τον φάκελο migrations. Το αρχείο αυτό θα περιλαμβάνει τις ενέργειες που θα πραγματοποιηθούν στην βάση. Το αρχείο αυτό παίρνει τις πληροφορίες από το αρχείο models.py. Ειδικότερα, η ροή ενέργειας του django περιβάλλοντος είναι να μετατρέπει τις κλάσεις του models.py σε tables στην βάση δεδομένων και τα πεδία των κλάσεων σε πεδία των tables. Κατά την εκτέλεση της δεύτερης εντολής δημιουργούνται οι απαραίτητες ενέργειες στην βάση δεδομένων.

Αφότου αναφέραμε πως θα δημιουργήσουμε τις στήλες και γενικότερα τα πλαίσια της βάσης δεδομένων, χρειάζεται να γίνει αναφορά στον τρόπο με τον οποίο θα εμπλουτίσουμε την βάση με πραγματικά στοιχεία. Τα δεδομένα αυτά μπορούν να προστεθούν με δύο τρόπους.

Ο πρώτος τρόπος εμπλουτισμού της βάσης είναι μέσω sql script. Ως SQL (Structured Query Language) script ορίζεται ένα κομμάτι κώδικα το οποίο είναι γραμμένο σε γλώσσα SQL και χρησιμοποιείται κυρίως για διαχείριση δεδομένων σε μια βάση. Αποτελεί μια εξαιρετικά απλή διαδικασία αφού το μοναδικό που χρειαζόμαστε είναι ένα εργαλείο σχεδίασης οπτικών βάσεων όπως είναι το SQL Workbench και οι SQL εντολές. Αφενός είναι μια απλή διαδικασία, αφετέρου αποτελεί μια χειροκίνητη, γεγονός που την κάνει επιρρεπή σε λάθη.

Ο δεύτερος τρόπος εμπλουτισμού είναι διαμέσου της εφαρμογής. Ειδικότερα, ο πιο ασφαλής τρόπος να προσθέσουμε πληροφορίες απαιτεί την χρήση των κλάσεων που δημιουργήσαμε. Περισσότερη ανάλυση θα γίνει στο επόμενο κεφάλαιο.

#	id	number	firstname	lastname	dob	height	weight	position	dominant_hand	nationality	experience	teamID_id
1	1	44	Bojan	Bogdanovic	1989-04-18	6-7	226	Small...	Right hand	Bosnia and H...	6	1
2	2	27	Rudy	Gobert	1992-06-26	7-1	258	Center	Right hand	France	7	1
3	3	23	Royce	O'Neale	1993-06-05	6-4	226	Small...	Right hand	Texas	3	1
4	4	20	Goerges	Niang	1993-06-17	6-7	230	Powe...	Right hand	Massachusetts	4	1
5	5	14	Derrick	Favors	1991-07-05	6-9	265	Powe...	Right hand	Georgia	10	1
6	6	45	Donovan	Mitchell	1996-09-07	6-1	215	Shoot...	Right hand	New York	3	1
7	7	0	Jordan	Clarkson	1992-06-07	6-4	194	Shoot...	Right hand	Florida	6	1
8	8	7	Joe	Ingles	1987-10-02	6-8	220	Small...	Left hand	Australia	6	1
9	9	10	Mike	Conley	1987-10-11	6-1	175	Point...	Left hand	Arkansas	13	1
10	10	22	Deandre	Ayton	1998-07-23	6-11	250	Center	Right	Bahamas	2	2
11	11	25	Mikal	Bridges	1996-08-30	6-6	209	Small...	Right	Pennsylvania	2	2
12	12	3	Chris	Paul	1985-05-06	6-0	175	Point ...	Right	North Carolina	15	2
13	13	1	Devin	Booker	1996-10-30	6-5	225	Shoot...	Right	Michigan	5	2

Εικόνα 5: Στιγμιότυπο Mysql βάσης δεδομένων

```

1 • INSERT INTO `basket_stats`.`run_playerstats` (`games`, `minutes_played`, `field_goals`, `field_goal_attempts`, `field_goal
2 VALUES ('68', '30.6', '5.2', '12.2', '.425', '2.5', '6.4', '.387', '2.7', '5.8', '.466', '2.9', '3.3', '.878', '0.
3 ('59', '31.1', '5.5', '8.2', '.671', '0.0', '0.1', '.000', '5.5', '8.1', '.675', '3.4', '5.5', '.618', '3.4', '16
4 ('59', '32.0', '2.4', '5.5', '.439', '1.5', '3.9', '.375', '0.9', '1.6', '.596', '0.6', '0.7', '.857', '1.2', '5.
5 ('60', '14.5', '2.2', '5.0', '.433', '1.5', '3.6', '.410', '0.7', '1.4', '.494', '0.2', '0.2', '1.000', '0.4', '2.
6 ('56', '15.5', '2.3', '3.7', '.617', '0.0', '0.0', '.000', '2.3', '3.6', '.623', '1.0', '1.3', '.740', '2.0', '3.
7 ('53', '33.4', '9.0', '20.6', '.438', '3.4', '8.7', '.386', '5.7', '11.9', '.476', '5.0', '6.0', '.845', '0.9', '3.
8 ('56', '26.0', '6.4', '15.1', '.422', '3.0', '8.7', '.349', '3.3', '6.4', '.522', '1.7', '1.8', '.922', '0.7', '3.
9 ('55', '27.4', '4.2', '8.1', '.517', '2.8', '5.8', '.488', '1.3', '2.2', '.593', '1.1', '1.3', '.857', '0.4', '3.
10 ('48', '29.8', '5.6', '12.6', '.444', '2.7', '6.7', '.409', '2.9', '5.9', '.482', '2.4', '2.7', '.863', '0.7', '2.
11 ('58', '22.8', '3.9', '8.6', '.452', '1.7', '3.9', '.441', '2.1', '4.6', '.461', '1.1', '1.3', '.805', '0.4', '2.
12 ('58', '28.4', '2.9', '6.4', '.458', '1.7', '4.2', '.406', '1.2', '2.2', '.559', '0.8', '0.9', '.846', '0.8', '4.
13 ('56', '19.9', '3.2', '6.6', '.477', '1.6', '3.6', '.460', '1.5', '3.1', '.497', '0.4', '0.5', '.828', '0.3', '2.
14 ('46', '34.4', '9.4', '18.2', '.516', '2.0', '5.0', '.393', '7.4', '13.2', '.562', '5.0', '5.7', '.878', '1.1', '5.
15 ('46', '33.7', '8.5', '17.8', '.479', '3.3', '7.6', '.429', '5.3', '10.2', '.517', '3.7', '4.2', '.891', '0.9', '5.
16 ('39', '23.6', '4.5', '8.9', '.507', '1.0', '2.8', '.352', '3.5', '6.1', '.577', '0.9', '1.2', '.822', '1.7', '4.

```

3.5 Στάδιο επεξεργασίας αθλητικών στατιστικών δεδομένων

Η επεξεργασία δεδομένων αποτελεί απαραίτητο κομμάτι σε κάθε έρευνα. Με δεδομένο ότι το παρόν project αποτελεί μια web εφαρμογή που δίνει δικαιώματα εγγραφής στον χρήστη, κρίνεται απαραίτητο να του δίνεται και το δικαίωμα της επεξεργασίας των δεδομένων αυτών. Ο όρος της επεξεργασίας ορίζεται από την απλή αλλαγή ονόματος της ομάδας του μέχρι την αλλαγή στατιστικών που αφορούν την κάθε ομάδα.

Για να επεξεργαστεί ο χρήστης τα δεδομένα που αναλογούν στην ομάδα του πρέπει να έχει τον αντίστοιχο ρόλο. Στα πλαίσια της εργασίας θεωρούμε πως κάθε υπεύθυνος ομάδας έχει κάνει εγγραφή στην εφαρμογή και έχει δημιουργήσει την ομάδα του, συνεπώς και το δικαίωμα να την επεξεργάζεται. Για να επεξεργαστεί ο αρμόδιος της προπονητικής ομάδας τα στατιστικά της ομάδας του ή των παικτών του χρειάζεται απλά να αναζητήσει το όνομα και να πατήσει την επιλογή “επεξεργασία στατιστικών” (edit stats). Η διαδικασία αυτή θα τον οδηγήσει σε μια προεπιλεγμένη φόρμα συμπλήρωσης όπου θα μπορεί να αλλάξει τις πληροφορίες που επιθυμεί εκείνος.

Εικόνα 7: Προεπιλεγμένη φόρμα συμπλήρωσης στοιχείων

Προγραμματιστικά αυτό επιτεύχθηκε με την χρήση django forms. Ειδικότερα, στο αρχείο forms.py δημιουργήσαμε κλάσεις στις οποίες κάνουμε αναφορά στα models που θέλουμε να αναφέρονται οι κλάσεις. Με αυτόν τον τρόπο δημιουργούμε μια αυτόματη διαδικασία που είναι ασφαλής από λάθη και απαιτεί λιγότερο κώδικα, συνεπώς και λιγότερη συντήρηση σε αυτόν. Στις κλάσεις αυτές ορίζουμε ποιά είναι τα πεδία που θέλουμε να περιέχουν οι φόρμες αλλά και τι μορφή θέλουμε να έχουν λχ στα πεδία που απαιτούν αριθμούς δεν είναι εφικτή η πληκτρολόγηση λέξης. Για την οπτικοποίηση των φορμών αυτών χρησιμοποιήθηκαν σελίδες html με χρήση css στοιχείων και javascript για την καλύτερη οπτικοποίηση των στοιχείων αυτών. Εκτενής αναφορά στην οπτικοποίηση θα γίνει σε ακόλουθο κεφάλαιο.

Ένα μεγάλο κομμάτι της επεξεργασίας των δεδομένων γίνεται αφότου οι πληροφορίες περάσουν το στάδιο των φορμών. Στα πλαίσια του django μηχανισμού, οι πληροφορίες αυτές μεταφέρονται στο αρχείο views.py. Στο αρχείο αυτό, δημιουργήσαμε μεθόδους οι οποίες είτε ελέγχουν τα δεδομένα που εισάγει ο χρήστης για τυχόν λάθη είτε συνδυάζουν τα στατιστικά αυτά για να δημιουργούν καινούργια. Χαρακτηριστικό παράδειγμα αποτελούν οι αμυντικές και επιθετικές διεκδικήσεις μιας ομάδας ενός παίκτη. Ο αρμόδιος υπεύθυνος της ομάδας συμπληρώνει τα δύο αυτά στατιστικά και παρατηρεί ότι αυτόματα έχει συμπληρωθεί και το πεδίο των συνολικών στατιστικών. Αντίστοιχη διαδικασία γίνεται και στο κομμάτι των εξειδικευμένων στατιστικών. Ο χρήστης πληκτρολογεί τον αριθμό των εύστοχων καλαθιών και των συνολικών προσπαθειών σε αυτά. Μέσω των μεθόδων στο αρχείο views υπολογίζεται το ποσοστό και αναγράφεται στην βάση. Τέλος αξίζει να σημειωθεί πως αν ο χρήστης πληκτρολογήσει κάποιο λάθος στην φόρμα ανακατευθύνεται σε αυτήν για να το διορθώσει.

```

def updatePlayerStats(request, id):
    try:
        playerStats = PlayerStats.objects.get(playerID=id)
        players = Players.objects.get(pk=id)
        form = PlayerStatsForm(request.POST or None, instance=playerStats)
    except Exception as e:
        print(e)
        form = PlayerStatsForm(request.POST or None)
        players = {}

    if form.is_valid():
        # games = form.cleaned_data.get('games')
        # minutes_played = form.cleaned_data.get('minutes_played')
        print('inside form')

```

Απόσπασμα κώδικα 3: Αρχείο `views.py` για την διαμόρφωση των στατικών του παίκτη

3.5 Οπτικοποίηση αποτελεσμάτων

Μετά την επεξεργασία των δεδομένων μέσω των μεθόδων απαιτείται και η κατάλληλη οπτικοποίησή τους. Για να επιτευχθεί αυτός ο σκοπός χρησιμοποιήσαμε σελίδες html με γλώσσα γραφής django template. Αυτός ο τρόπος γραφής ορίζεται με την χρήση κάποιων ειδικών html tags τα οποία αναγνωρίζονται από το πλαίσιο της django. Ωστόσο η django αναγνωρίζει τις html σελίδες μόνο αν βρίσκονται σε μια συγκεκριμένη θέση. Για να ορίσουμε την θέση αυτή πρέπει να βάλουμε στο αρχείο `settings.py`, το κατάλληλο μονοπάτι κάτω από το οποίο θα βρίσκονται τα templates. Στο αρχείο αυτό, θα παρατηρήσουμε μια μεταβλητή `templates` και ένα πεδίο της που ονομάζεται `"DIRS"`. Στο πεδίο αυτό πληκτρολογούμε την ακόλουθη εντολή `[os.path.join(BASE_DIR, 'templates')]`. Αυτή η εντολή πρακτικά σημαίνει πως οι html σελίδες θα βρίσκονται σε έναν φάκελο `templates`, ο οποίος βρίσκεται κάτω από το `base directory`, το οποίο αποτελεί τον φάκελο του project μας.

```

TEMPLATES = [
    {
        'BACKEND': 'django.template.backends.django.DjangoTemplates',
        'DIRS': [os.path.join(BASE_DIR, 'templates')],
        'APP_DIRS': True,
        'OPTIONS': {

```

Απόσπασμα κώδικα 4: Ρυθμίσεις `settings.py` για την χρήση `templates`

```

<div class="select">
    <select class="select-box" name="teamID" id="teamID">
        <option selected disabled>Choose your team's name</option>
        {% for x,y in form.fields.teamID.choices %}-->
            <div class="option">
                <option value="{{ x }}">{{ y }}</option>
            </div>
        {% endfor %}
    </select>

```

Απόσπασμα κώδικα 5: `html` κώδικας με
χρήση `django template tags`

4

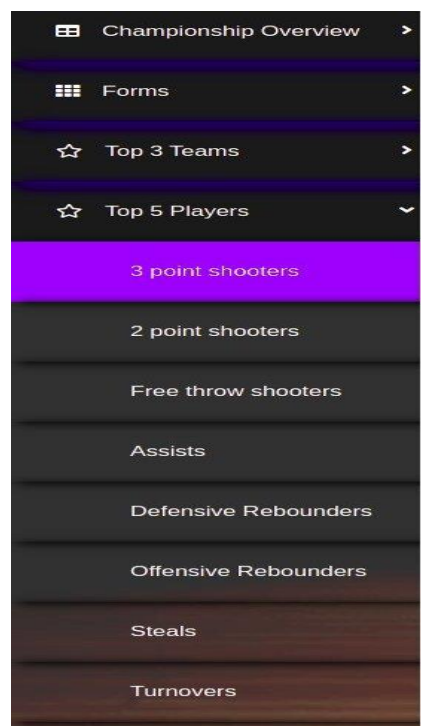
Πειραματικά αποτελέσματα

Σκοπός του κεφαλαίου αυτού είναι η παράθεση και επεξηγηματική ανάλυση των αποτελεσμάτων των προαναφερόμενων διαδικασιών για την αποτύπωση διαγραμμάτων με στατιστικά αποτελέσματα. Παρακάτω, γίνεται αναλυτική παρουσίαση των πειραματικών αποτελεσμάτων μας σε συνδυασμό με ταυτόχρονη διευκρίνιση λεπτομερειών. Σε αυτό το σημείο, αξίζει να δοθεί έμφαση στο μεγάλο εύρος των δυνατοτήτων εξέλιξης των παραπάνω διαδικασιών με σκοπό την ύπαρξη πιο εύστοχων και λεπτομερών αποτελεσμάτων. Η δυσκολία των παραπάνω τεχνικών σε συνδυασμό με την δυσκολία συγκέντρωσης δεδομένων καθιστούν σαφές το γεγονός της ύπαρξης περιθωρίου βελτίωσης των διαδικασιών και ακολούθως την αποτύπωση ακριβέστερων αποτελεσμάτων.

4.1 Παρουσίαση αποτελεσμάτων

Τα δεδομένα που εξήγαμε κατηγοριοποιούνται στους αθλητές αλλά τις ομάδες οι οποίες διέπουν σε κάθε κατηγορία. Ειδικότερα, από τις δέκα ομάδες που είχαμε ως δείγμα δεδομένων ξεχωρίσαμε τις τρεις οι οποίες είναι οι πιο αποτελεσματικές στην κατηγορία των πόντων, διεκδικήσεων, εύστοχων προσπάθειών κτλ. Για τους αθλητές είχαμε ένα δείγμα περίπου εκατό ατόμων οπότε έχουμε μια καλύτερη εικόνα των συνολικών τους επιδόσεων και έχουμε μεγαλύτερη δυνατότητα να τους συγκρίνουμε μεταξύ τους. Στο κομμάτι των αθλητών επιλέξαμε να παρουσιάσουμε τους πέντε κορυφαίους σε κάθε κατηγορία.

Ο λόγος που επιλέξαμε να διαχωρίσουμε αυτές τις κατηγορίες είναι επειδή αποτελούν σημαντικά στοιχεία και προκύπτουν αξιοσημείωτα συμπεράσματα. Πιο συγκεκριμένα, είναι εξαιρετικά σημαντική η πληροφορία να γνωρίζεις τον αριθμό των πόντων που σκοράρει μια ομάδα κατά μέσο όρο. Αντίστοιχα, είναι εξίσου βοηθητικό να γνωρίζεις αν μια ομάδα σκοράρει κυρίως με τρίποντα ή αν είναι κακή σε αυτό. Όλα τα προαναφερθέντα σε συνδυασμό με τον ανθρώπινο παράγοντα που θα πράττει σύμφωνα με αυτά αποτελούν μια καινοτόμο δράση στον αθλητικό τομέα.



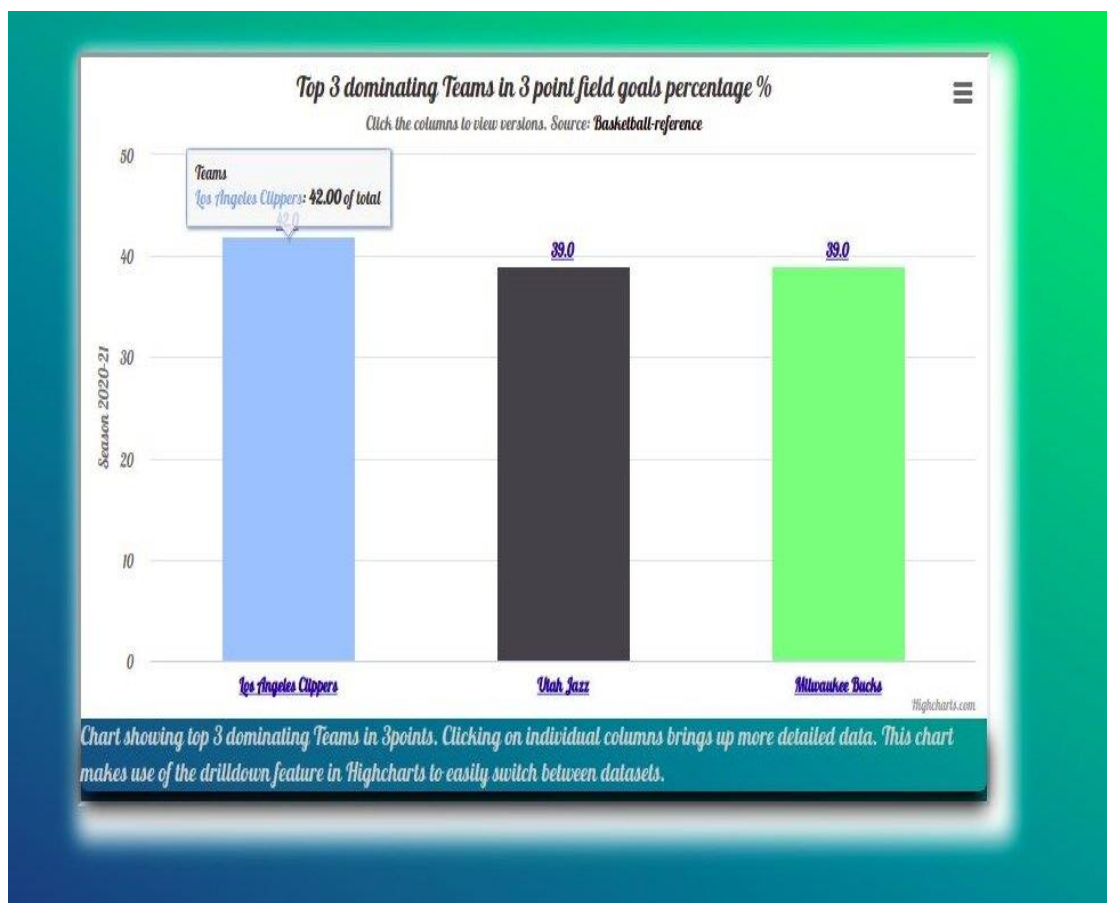
Εικόνα 8: Μενού UI για την κατηγοριοποίηση

4.2 Οπτικοποίηση συμπερασμάτων

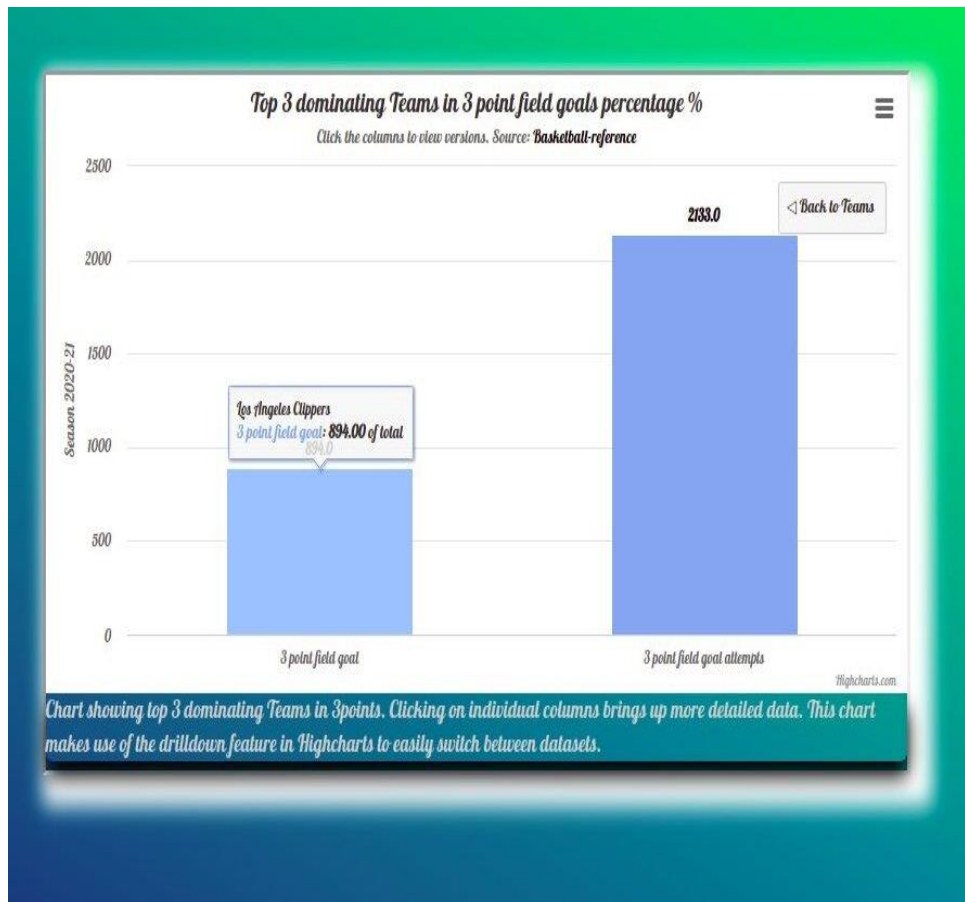
Για την καλύτερη οπτικοποίηση των συμπερασμάτων επιλέξαμε να γίνουν μέσω διαγραμμάτων. Με αυτόν τον τρόπο είναι πιο ευχάριστα για τον επισκέπτη αλλά αποτυπώνουν την πληροφορία με καλύτερο τρόπο αφού δίνουν περισσότερες δυνατότητες αξιοποίησης. Τα διαγράμματα που θεωρήσαμε πως κάλυπταν

καλύτερα τους στόχους διακρίνονται σε τρία είδη: διαγράμματα μπάρας (bar charts), διαγράμματα σφαίρας (bullet charts) και διαγράμματα πίτας (pie charts).

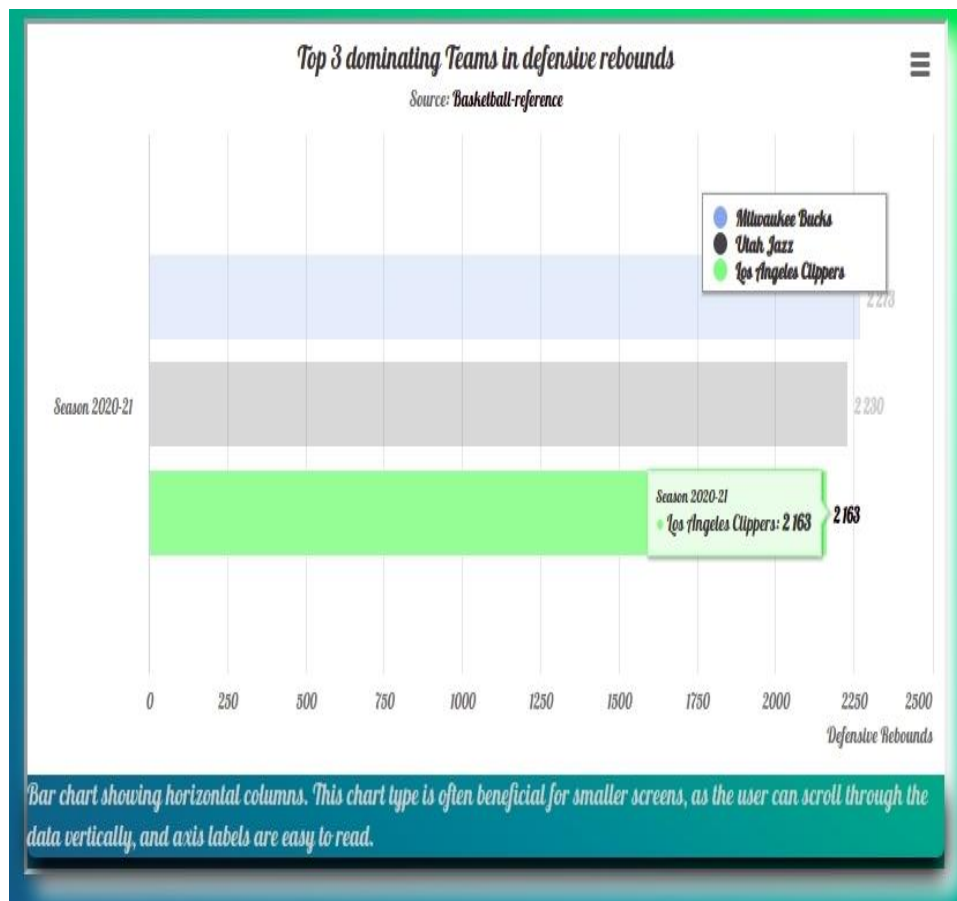
- Τα διαγράμματα μπάρας χρησιμοποιήθηκαν στις εξής κατηγορίες, 3 point chart, 2 point chart, defensive rebounds, offensive rebounds και points chart. Ο λόγος που επιλέξαμε την συγκεκριμένη κατηγορία διαγραμμάτων οφείλεται στην δομή τους. Ειδικότερα, στο διάγραμμα τριών πόντων βλέπουμε να απεικονίζεται το όνομα της ομάδας και το ποσοστό της στα σουτ τριών πόντων. Σε περίπτωση που επιθυμούμε παραπάνω στατιστικά, επιλέγουμε την αντίστοιχη μπάρα και βλέπουμε ένα νέο διάγραμμα μπάρας το οποίο αναφέρει πόσες εύστοχες προσπάθειες έχουν και πόσες συνολικές προσπάθειες. Από την διαίρεση των δυο αυτών αριθμών προέκυψε το αρχικό ποσοστό που είδαμε. Αντίστοιχη λογική αποτελούν και οι λοιπές κατηγορίες. Ιδιαίτερη σημείωση θέλει το διάγραμμα των πόντων αφού αποτελεί πολύ σημαντικό στατιστικό στον τομέα του αθλητισμού. Στο συγκεκριμένο διάγραμμα παρατηρούμε πως οι γραμμές είναι συρόμενες (draggable lines), χαρακτηριστικό που δίνει στον χρήστη δυνατότητα πιο μελετημένου σχεδιασμού. Επιπλέον χαρακτηριστικό αποτελεί ο δείκτης μέσου όρου, αφού τονίζει τον μέσο όρο των πόντων που βάζουν οι ομάδες. Σε περίπτωση που μια ομάδα βρίσκεται πάνω από τον μέσο όρο εμφανίζεται με πράσινο χρώμα ενώ αν βρίσκεται κάτω με κίτρινο.



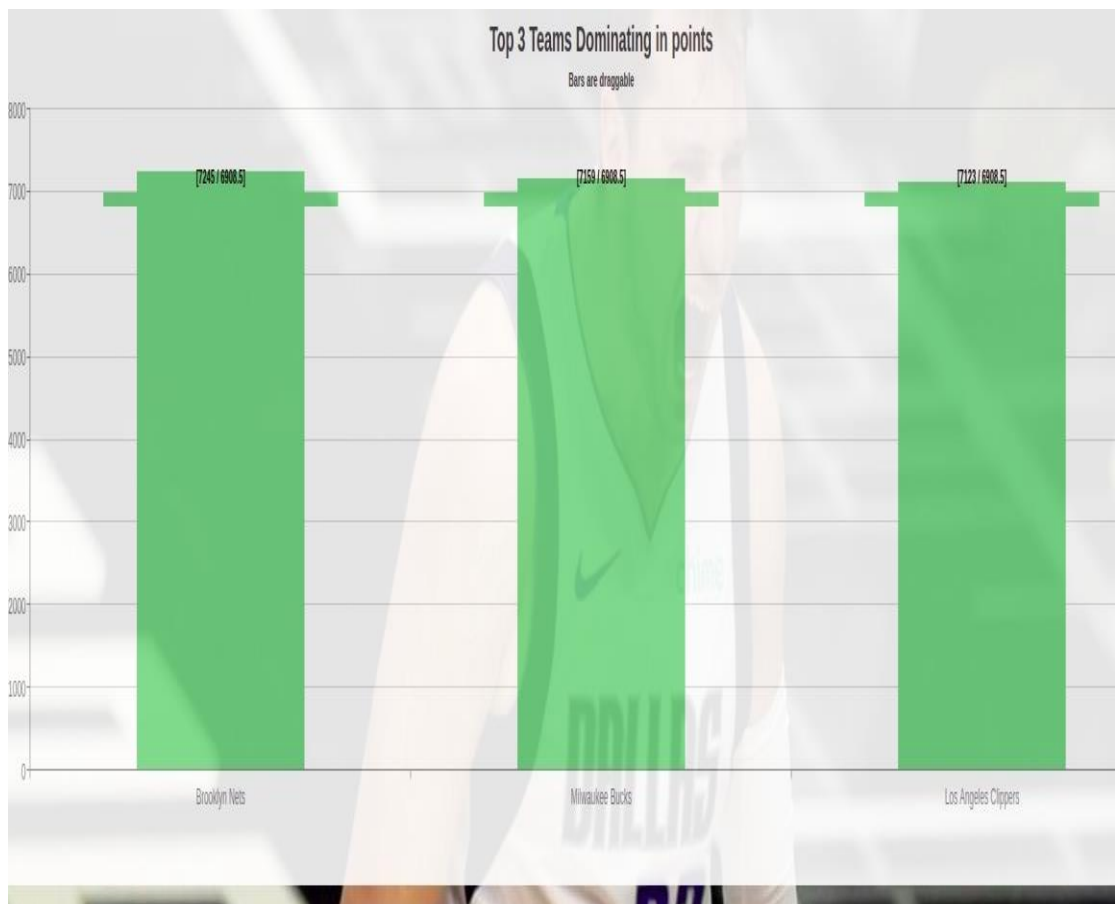
Διάγραμμα 1: Κορυφαίες ομάδες σε ποσοστό τριπόντων



Διάγραμμα 2: Διευρυμένο διάγραμμα που απεικονίζει εύστοχες και συνολικές προσπάθειες

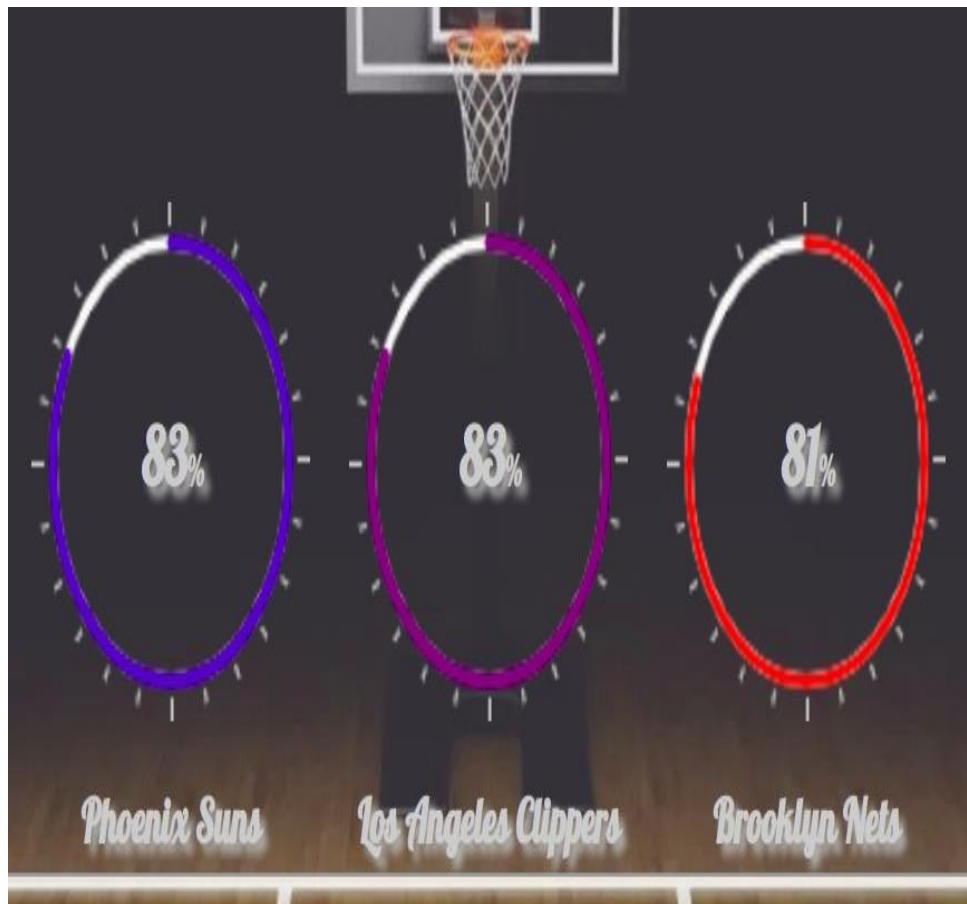


Διάγραμμα 3: Διάγραμμα μπάρας που απεικονίζει τις τρεις κορυφαίες ομάδες σε αμυντικών διεκδικήσεων



Διάγραμμα 4: Διάγραμμα μπάρας που απεικονίζει τις τρεις κορυφαίες ομάδες σε πόντους

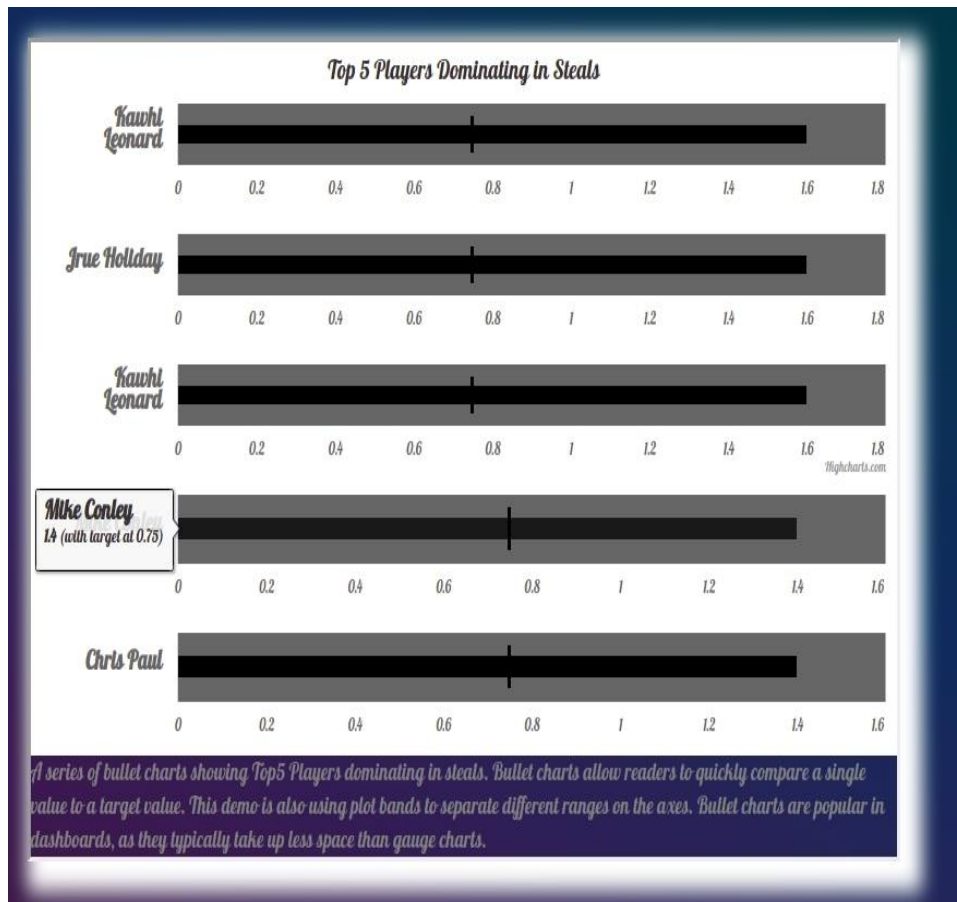
- Τα διαγράμματα πίτας αποτελούν μια πιο απλή μορφή διαγράμματος, με σημαντική αξία ωστόσο. Το πεδίο που παρουσιάστηκε με αυτόν τον τρόπο είναι το ποσοστό των ελευθέρων βολών. Με δεδομένη την απλότητα του διαγράμματος αυτού, δόθηκε ιδιαίτερη έμφαση στην διαδραστικότητα. Αφότου επιλέξουμε το όνομα ενός παίκτη μεταφερόμαστε στην σελίδα που έχει τα πλήρη στατιστικά του.



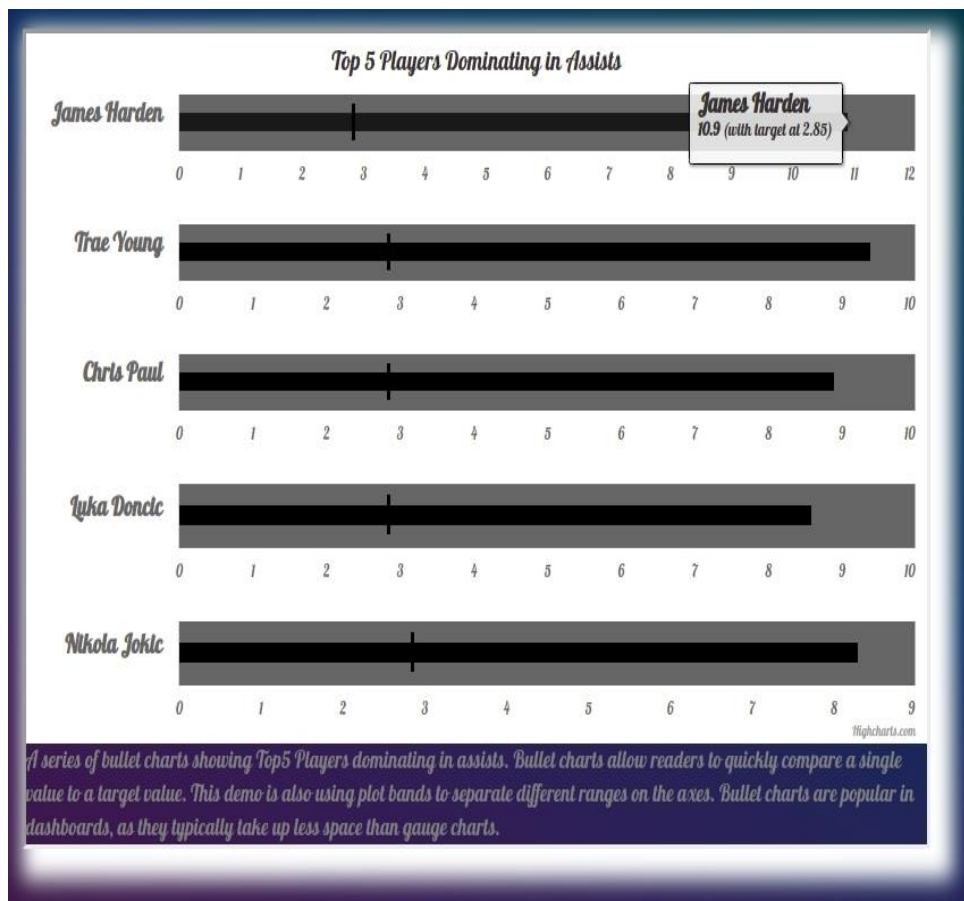
Διάγραμμα 5: Διάγραμμα πίτας που απεικονίζει τις τρεις κορυφαίες ομάδες σε εύστοχες βολές.

- Τελευταία κατηγορία διαγραμμάτων αποτελούν τα διαγράμματα σφαίρας (bullet diagram). Αποτελούν σχεδιαγράμματα με λίγα τεχνικά χαρακτηριστικά(λχ δεν κάνουν εξαγωγή πληροφορίας), ωστόσο περιέχουν μεγάλο όγκο πληροφορίας. Ειδικότερα, τέτοια διαγράμματα χρησιμοποιήθηκαν για την προβολή στατιστικών στα πεδία των λαθών(turnovers), πόντων(points), τελικών πασών(assists) και κλεψιμάτων(steals). Η αιτία επιλογής του συγκεκριμένου σχεδιαγράμματος οφείλεται στην εύκολη σύγκριση τιμών. Ειδικότερα, προβάλλεται η τιμή ενός πεδίου στα στατιστικά ενός αθλητή με την χρήση μιας απλής μπάρας και με μια δεύτερη μπάρα προβάλλεται ο

μέσος όρος του πεδίου αυτού. Η γρήγορη σύγκριση τιμών εξυπηρετεί για λόγους χρησιμότητας αφού μπορούμε γρήγορα να έχουμε την σύγκριση τιμών – μέσου όρου.



Διάγραμμα 6: Διάγραμμα σφαίρας που απεικονίζει τους πέντε κορυφαίους παίκτες σε κλεψίματα.



Διάγραμμα 7: Διάγραμμα σφαίρας που απεικονίζει τους πέντε κορυφαίους παίκτες σε τελικές πάσες.

Προγραμματιστικά τα παραπάνω υλοποιήθηκαν με την χρήση γλώσσας προγραμματισμού HTML συνδυαστικά με JavaScript. Η HTML ή Hypertext Markup Language είναι μια τυπική επιλογή που χρησιμοποιείται για την ανάπτυξη ιστοτόπων, ιστοσελίδων και εφαρμογών που βασίζονται στον Ιστό(web). Κάποια από τα θετικά της HTML είναι πως αποτελεί μια ελαφριά γλώσσα προγραμματισμού, υποστηρίζεται από όλους τους περιηγητές ιστού (web browsers) και είναι εύκολη για τον προγραμματιστή. Η JavaScript είναι μια δυναμική γλώσσα προγραμματισμού υπολογιστών. Είναι ελαφριά και χρησιμοποιείται συνήθως ως μέρος των ιστοσελίδων, του οποίου η εφαρμογή επιτρέπει σενάριο πελάτη να αλληλεπιδρά με έναν χρήστη και να δημιουργεί δυναμικές σελίδες. Είναι μια ερμηνευμένη γλώσσα προγραμματισμού με αντικειμενοστραφείς δυνατότητες.

Στην παρούσα έρευνα, χρησιμοποιήθηκε πλήθος κώδικα σε JavaScript και jQuery βιβλιοθήκης για να δημιουργηθούν δυναμικά σχέδια και να γίνει πιο ενδιαφέρουσα η εμπειρία για τον χρήστη. Επίσης, δημιουργήθηκαν δυναμικά διαγράμματα (charts) για την

καλύτερη αποτύπωση και ερμηνευση συμπερασμάτων. Η πλειοψηφία τους προήλθε από εισαγωγή από τον ιστότοπο *highcharts* (link <https://www.highcharts.com/>) και έπειτα έγινε η επεξεργασία τους για την κάλυψη των αναγκών της παρούσας εργασίας.

5

Επίλογος

5.1 Σύνοψη και συμπεράσματα

Ανακεφαλαιώνοντας, στη παρούσα μελέτη ασχοληθήκαμε με τη ανάλυση αθλητικών δεδομένων και η έρευνα μας χωρίστηκε σε τρία βασικά στάδια. Αρχικά, πραγματοποιήθηκε η συλλογή και εξόρυξη δεδομένων και πιο συγκεκριμένα η εύρεση των στατιστικών στοιχείων των ομάδων και των αθλητών του NBA.

Για τη διαδικασία αυτή χρησιμοποιήθηκε όπως έχει ήδη προαναφερθεί η MySQL βάση, στην οποία εκτελέστηκε το query αναζήτησης των κορυφαίων αθλητών και ομάδων σε κάθε κατηγορία επιθετικά και αμυντικά. Στον τομέα των αθλητών χρησιμοποιήθηκαν επιπλέον κριτήρια για να εξαιρεθούν οι αθλητές που είχαν μικρό χρόνο συμμετοχής, συνεπώς μη αντιπροσωπευτικά χαρακτηριστικά.

Στο στάδιο αυτό, παρουσιάστηκαν αρκετά σφάλματα κατά την επιλογή των δεδομένων. Αρχικά δεν υπάρχει μια αρμόδια ιστοσελίδα που να μεταφέρει τις πληροφορίες αυτές αξιόπιστα με την μορφή API. Αυτό έχει ως αποτέλεσμα οι πληροφορίες να περαστούν μέσω queries στην MySQL βάση, γεγονός που μπορεί να επιφέρει αρκετά λάθη αφού εμπλέκεται ο ανθρώπινος παράγοντας σε μεγάλο βαθμό. Ένα πρόβλημα που προέκυψε εξίσου αφορά ποιες πληροφορίες κρίνονται απαραίτητες για την έρευνα αφού αρκετές ιστοσελίδες περιείχαν εξειδικευμένα δεδομένα τα οποία δεν αναλύουν πώς μπορούν να συνδράμουν σε μια έρευνα.

Ακολούθως, η μελέτη συνεχίστηκε με τη διαχείριση των αθλητικών δεδομένων. Αναλυτικότερα, πραγματοποιήθηκε διαχωρισμός και ομαδοποίηση των

δεδομένων με την ανάπτυξη κώδικα σε Python και ειδικότερα σε Django framework. Αρχικά, δημιουργήθηκαν κλάσεις για την καλύτερη κατηγοριοποίηση των δεδομένων. Έπειτα προστέθηκε βασική λειτουργία μιας web εφαρμογής για την καλύτερη αξιοποίηση των δεδομένων.

Στο στάδιο διαχωρισμού και ομαδοποίησης εξάγαμε χρήσιμα συμπεράσματα. Ειδικότερα, διαχωρίσαμε της ομάδες που ξεχωρίζουν επιθετικά με την μορφή ευστοχίας(τρίποντα, δίποντα) αλλά και αμυντικά (αμυντικές διεκδικήσεις, κλεψίματα). Για τον διαχωρισμό αυτόν χρησιμοποιήσαμε κριτήρια όπως το ποσοστό ευστοχίας στα queries που δημιουργήθηκαν. Για την καλύτερη εξαγωγή συμπερασμάτων δεν συμπεριλάβαμε τους παίκτες που έχουν λιγότερες από πενήντα οκτώ (58) συμμετοχές σε αγώνες, δηλαδή το εβδομήντα(70) τοις εκατό των συνολικών αγώνων. Το τελευταίο κριτήριο αποτελεί πραγματικό κριτήριο των στατιστικών του NBA, αφού θεωρείται πως οποιοσδήποτε αθλητής δεν συμπληρώνει το εβδομήντα τοις εκατό δεν μπορεί να διεκδικήσει κανέναν ατομικό τίτλο λχ πολυτιμότερου παίκτη (MVP).

Ολοκληρώνοντας, η μελέτη συνεχίστηκε με την οπτικοποίηση των παραπάνω συμπερασμάτων. Η οπτικοποίηση έλαβε την μορφή γραφημάτων, σχημάτων δηλαδή που είναι πιο κατανοητά και περικλείουν μεγάλο όγκο δεδομένων. Από το πλήθος διαγραμμάτων επιλέχθηκαν να χρησιμοποιηθούν τρία είδη γραφημάτων, σφαίρας, μπάρας και πίτας. Τα διαγράμματα πίτας αποτελούν την πιο απλή μορφή των τριών και γι αυτόν τον λόγο χρησιμοποιήθηκαν για την απεικόνιση ποσοστών στις ελεύθερες βολές τόσο των ομάδων όσο και των αθλητών. Τα διαγράμματα μπάρας δίνουν περισσότερες πληροφορίες και χρησιμοποιήθηκαν σε περισσότερα πεδία. Με την χρήση τους αυτή, μπορούμε να κάνουμε εύκολη σύγκριση τιμών μεταξύ τους συνεπώς πιο γρήγορα αποτελέσματα. Επιπλέον με την χρήση της JavaScript γλώσσας προγραμματισμού που χρησιμοποιήθηκε μπορούμε να εξάγουμε το αντίστοιχο διάγραμμα, να αποτυπώσουμε τις τιμές του διαγράμματος σε μορφή πίνακα ή και να το μετατρέψουμε σε εικόνα ή απλό αρχείο διαχωρισμένο με κόμματα (CSV file). Τέλος, τα διαγράμματα σφαίρας αποτυπώνουν σημαντικές πληροφορίες όσον αφορά την σύγκριση μιας ομάδας με τον μέσο όρο και όχι τόσο

μεταξύ των ομάδων μεταξύ τους. Με τον σκοπό αυτό, απεικόνισαν πεδία που μας ενδιαφέρει η σχέση τιμής – μέσου όρου όπως το πεδίο των πόντων.

Όσον αφορά την έρευνά μας, η προσπάθεια αποτύπωσης των κορυφαίων ομάδων και αθλητών σε κάθε κατηγορία παρουσίασε πλούσια συμπεράσματα. Πρώτα από όλα, είναι ευδιάκριτη η υπεροχή της ομάδας σε άμυνα ή επίθεση συγκριτικά με άλλη ομάδα.

Με τον τρόπο αυτόν, μπορούν να γίνουν εκτιμήσεις όσον αφορά τους πόντους που θα δεχθεί μια ομάδα στην άμυνα και πόσους πόντους θα σκοράρει στην επίθεση.

Έπειτα υπάρχει διαφάνεια μεταξύ των πεδίων οπότε είναι εύκολο να διακριθεί αν μια ομάδα υστερεί σε ένα πεδίο λχ ποσοστό διπόντων. Συνεπώς είναι πιο εύκολο τόσο για την ομάδα όσο και για τον αθλητή να εστιάσει στον τομέα αυτόν.

Τέλος, σημαντικό πρόβλημα στο οποίο η έρευνα έχει την δυνατότητα να προσφέρει λύση, αποτελεί η λήψη αποφάσεων της προπονητικής ομάδας. Με την ανάλυση της κάθε ομάδας επιθετικά και αμυντικά είναι πιο εύκολη η αντιμετώπιση της (scouting), αλλά και πιο άμεση η βελτίωση της ίδιας της ομάδας. Επιπλέον με τα εξειδικευμένα στατιστικά των παικτών είναι πιο εύκολη η πρόβλεψη της απόδοσης ενός παίκτη συνεπώς και την αντιμετώπισή του.

5.2 Μελλοντικές επεκτάσεις

Αναφορικά με τη μελλοντική εργασία, αποτελεί ενδιαφέρον αρχικά η εφαρμογή των παραπάνω συμπερασμάτων στην δημιουργία ενός αλγορίθμου ο οποίος θα μπορεί να συντάσσει την καλύτερη πεντάδα αθλητών με την χρήση των ακριβών στατιστικών τους. Αυτό θα αποτελεί μια ασφαλή λύση στις ομάδες που δυσκολεύονται να δημιουργήσουν μια αποδοτική πεντάδα αθλητών.

Επιπλέον θα ήταν αρκετά ωφέλιμη η δημιουργία API (Application Programming Interface) τα οποία θα επικοινωνούν με την εφαρμογή και θα παρέχουν νέες πληροφορίες όσων αφορά τις μεταγραφές των αθλητών αλλά και τα ποσοστά τους σε κάθε αγώνα. Αυτό θα έχει ως συνέπεια, η εφαρμογή να γίνει

περισσότερο αυτοματοποιημένη και να απαιτεί λιγότερη συμμετοχή από τον διαχειριστή. Επιπλέον θα έχει μικρότερο περιθώριο λάθους εφόσον θα είναι πιο αυτοματοποιημένη διαδικασία και η συντήρηση της εφαρμογής θα είναι πιο εύκολη.

Τέλος, μια εξίσου σημαντική επέκταση είναι η κάλυψη στατιστικών ανά αγώνα. Με αυτόν τον τρόπο θα υπάρχει μια μεγαλύτερη βάση δεδομένων από την οποία θα γίνονται οι συγκρίσεις. Επιπλέον, θα είναι εφικτή η απεικόνιση διακύμανσης της απόδοσης του αθλητή ή της ομάδας σε βάθος χρόνου. Με αυτόν τον τρόπο θα προκύψουν νέα συμπεράσματα, συνεπώς νέο υλικό για μελέτη.

Βιβλιογραφία

- 1 Toshihiro Tani, Hung-Hsuan Huang, Kyoji Kawagoe. Sports Play Visualization System Using Trajectory Mining Method (2015)
- 2 Toshihiro Tani, Hung-Hsuan Huang, Kyoji Kawagoe. Sports Play Visualization System for American Football (2015)
- 3 JuhaOksanen, Cecilia Bergman, Jani Sainio, Jan Westerholm. Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data. (2015)
- 4 Laura Ferrari, Marco Mamei. Identifying and understanding urban sport areas using Nokia Sports Tracker (2013)
- 5 Fosca Giannotti, Mirco Nanni, Dino Pedreschi, Fabio Pinelli, Chiara Renso, Salvatore Rinzivillo, Roberto Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data (2011)
- 6 Hoyoung Jeung, Man Lung Yiu, Christian S. Jensen. Trajectory Pattern Mining
- 7 Long Sha, Patrick Lucey, Yisong Yue, Xinyu Wei, Jennifer Hobbs, Charlie Rohlf, Sridha Sridharan. Interactive Sports Analytics: An Intelligent Interface for Utilizing Trajectories for Interactive Sports Play Retrieval and Analytics (2018)
- 8 Tutorialspoint. Advantages and Disadvantages of JavaScript
- 9 Youtube, Corey Schafer. Python Django Tutorial
- 10 Youtube, Very Academy. Learn Django Class-Based Views
- 11 Youtube, Professional Cipher. How to connect mySQL database with Django Project
- 12 Basketball reference <https://www.basketball-reference.com>