**ASSIGNMENT 2 FRONT SHEET**

|  |  |  |  |
| --- | --- | --- | --- |
| **Qualification** | **BTEC Level 5 HND Diploma in Computing** | | |
| **Unit number and title** | Unit 14: Business Intelligence | | |
| **Submission date** |  | **Date Received 1st submission** |  |
| **Re-submission Date** |  | **Date Received 2nd submission** |  |
| **Student Name** | Pham Quang Minh | **Student ID** | GCh18572 |
| **Class** | GCH0714 | **Assessor name** | Doan Trung Tung |
| **Student declaration**  I certify that the assignment submission is entirely my own work and I fully understand the consequences of plagiarism. I understand that making a false declaration is a form of malpractice. | | | |
|  |  | **Student’s signature** | Minh |

**Grading grid**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| P3 | P4 | P5 | P6 | M3 | M4 | D3 | D4 |
|  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **❒ Summative Feedback: ❒ Resubmission Feedback:** | | |
| **Grade:** | **Assessor Signature:** | **Date:** |
| **IV Signature:** | | |

Contents

[Introduction: 5](#_Toc60345415)

[General about BI 5](#_Toc60345416)

[What is BI 5](#_Toc60345417)

[Real examples of how to apply BI on business 6](#_Toc60345418)

[BI techniques: 7](#_Toc60345419)

[BI tools: 9](#_Toc60345420)

[Tableau 9](#_Toc60345421)

[Python 10](#_Toc60345422)

[Excel 11](#_Toc60345423)

[Explain dataset 12](#_Toc60345424)

[Explain pre-process steps on dataset 13](#_Toc60345425)

[Design Dashboards: 19](#_Toc60345426)

[It is possible to choose the first player, because his value is absolutely superior to the rest, his salary is not in the top. However, if the economy is not enough, the team can choose other players with much less money 28](#_Toc60345427)

[Discuss the legal issues involved in exploiting user data for business intelligence 28](#_Toc60345428)

[Discuss how business intelligence tools can contribute to effective decision-making 29](#_Toc60345429)

[Feedback/comment 31](#_Toc60345430)

[Bibliography 34](#_Toc60345431)

[Figure 1: tableau 10](#_Toc60345432)

[Figure 2: Python 11](#_Toc60345433)

[Figure 3: Excel 12](#_Toc60345434)

[Figure 4: code open 13](#_Toc60345435)

[Figure 5: Code delete columns 13](#_Toc60345436)

[Figure 6: code Change symbols to letters with the same meaning 14](#_Toc60345437)

[Figure 7: code Correct value 15](#_Toc60345438)

[Figure 8: code Reputation of wage 15](#_Toc60345439)

[Figure 9: code Edit Height 16](#_Toc60345440)

[Figure 10: code Fix weight 16](#_Toc60345441)

[Figure 11: code Indicates the position of each player on the field 17](#_Toc60345442)

[Figure 12: code fix Release Clause 18](#_Toc60345443)

[Figure 13: code Split the column in two 18](#_Toc60345444)

[Figure 14: code Export the cleaned data to a new excel file 19](#_Toc60345445)

[Figure 15: Dashboard 1 19](#_Toc60345446)

[Figure 16: Aim to find out which players have low stamina and speed on the team 20](#_Toc60345447)

[Figure 17:After I discovered those players, I looked at where they played, and what age they were 21](#_Toc60345448)

[Figure 18: After I have identified the players with low stats, I find out if they are still performing well in the stats 21](#_Toc60345449)

[Figure 19: Interaction on dashboard 22](#_Toc60345450)

[Figure 20: Dashboard 2 24](#_Toc60345451)

[Figure 21: Contain charts 25](#_Toc60345452)

[Figure 22: wage 26](#_Toc60345453)

[Figure 23: Age 26](#_Toc60345454)

[Figure 24: Interaction on dashboard 27](#_Toc60345455)

[Figure 25: Dashboard 29](#_Toc60345456)

[Figure 26: Highlight data 30](#_Toc60345457)

[Figure 27: Highlight Data 31](#_Toc60345458)

[Figure 28: datasets code against Python 31](#_Toc60345459)

[Figure 29: comments on our code 32](#_Toc60345460)

[Figure 30: rate chart and dashboard 32](#_Toc60345461)

[Figure 31: suggestions on our chart and dashboard 33](#_Toc60345462)

[Figure 32:our presentation 33](#_Toc60345463)

# Introduction:

Our next mission is to demonstrate to the board the company's ability to apply business intelligence to the company's current business processes. To demonstrate BI, we prepared a presentation on BI and related tools & techniques and a demonstration on real company data set. First we will explain what the general concept of BI is. Introduce some tools / techniques for BI and their application in general. Next we show a dataset extracted from company business processes. Explain the data set. Shows how we preprocess the data for later analysis, explaining each step and its purpose. Design dashboards to display our analytics on pre-processed data. Clearly explain the purpose of the dashboard and chart. Gather feedback and comments from users to gauge how responsive a user or business is required, and customize it necessary for future use. Business intelligence can contribute to effective decision making as well as legal issues related to user data mining for business intelligence operations. Research specific examples of organizations that use BI tools to enhance or improve their business and evaluate how BI tools can be used to broaden their target audience and be more competitive. market. presents our views on how business intelligence tools can contribute to effective decision-making as well as legal issues related to user data mining for business intelligence. Study specific examples of organizations using BI tools to enhance or improve their business and evaluate how BI tools can be used to broaden their target audience and be more competitive on market. presents our views on how business intelligence tools can contribute to effective decision-making as well as legal issues related to user data mining for business intelligence. Research specific examples of organizations that use BI tools to enhance or improve their business and evaluate how BI tools can be used to broaden their target audience and be more competitive market.

# General about BI

## What is BI

* BI (Business Intelligence) is a set of processes, architectures, and technologies that convert raw data into meaningful information that drives profitable business actions. It is a suite of software and services for converting data into actionable intelligence and knowledge.
* BI has a direct influence on the strategic, tactical, and operational business decisions of the organization. BI supports fact-based decision-making using historical data rather than assumptions and feelings.
* BI tools perform data analysis and generate reports, summaries, dashboards, maps, graphs, and charts to give users insight into the nature of the business.

(GURU99, n.d.)

## Real examples of how to apply BI on business

|  |  |
| --- | --- |
| Information | Business intelligence |
| Add information of players  Player database  Change information of players  Add diets for players  Nutrition specialist for football players  Change diets for players | Add new line for player  Change information (delete, update)  Data Warehouse  Add new line diet for player  Quality indicators of players after using the menu that were requested |

-in the diet management system of online players. The players' information is included in the database from expert player management and player rating.

* Add information of players
* Change information of players

Correspondingly, in a Business Intelligence system query will be executed for the player be it the addition of new player lines or changing player information. the aim is to regularly update the player's status, to best evaluate his condition and characteristics. from there, it helps doctors and experts to create a daily menu for each player.

-In an football player database of nutrition specialist system query that could be executed

* Add diets for players
* Change diets for players

Correspondingly, in the executable BI system query would be the Player Quality Index after using the requested menu.

## BI techniques:

Collection techniques:

1.Cleansing

Data cleaning is the process of making sure it is accurate, reliable, and accessible. To avoid similar errors from occurring, you can clean up the data by finding errors or corrupt, correct or delete them, or process the data manually if needed. (Leo Gimenez, 2020).

* Data cleaning is looking for ways to improve the accuracy of the data set without actually deleting the information.
* Data cleansing is known as a fundamental aspect of the fundamentals of data science, as it plays an important role in the process of analyzing and giving accurate feedback.
* Business organizations need to pay attention to the accuracy of data for information to be used effectively and meaningfully.

2. Labeling

Labeling is the method by which data samples are detected and labeled. You ensure your in-app dashboards and reports look the same as the rest of your applications when you blank the BI software. By maintaining brand continuity and a smooth user experience, successful white-labeling increases user adoption of BI and embedded analytics. (Logi Analytics, 2020).

* The labeling is done manually and automatically.
* Data labeling provides labels to categorize input and output data, providing a learning base for potential data processing.

Analysis techniques :

1. Reporting: One area where BI tools help business users is design, schedule and generate results, report sales, collate and save. Reports generated by BI tools efficiently collect and present data to help with the management, planning, and decision-making processes.
2. Dashboard: Dashboard is a major feature of BI. Dashboards serve the need to process and organize details into an easy to understand format. You can instantly get valuable insights that you can use to make data-driven business decisions by looking at the dashboard. (Bergen Adair, 2019).

OLAP(on-line analytical processing): is a high-speed multidimensional research software for large amount of data from a data center, data center or some other centralized, aggregated data warehouse. OLAP extracts and reorganizes information from several relational data sets into a multidimensional format that enables very fast processing and very in-depth analysis (IBM Cloud Education, 2020).

Data mining: Data mining is a method used in large databases to explore patterns and is also related to database systems, statistics, and machine learning to identify these patterns. Data mining is an integrated method for both data management and data preprocessing, as it ensures the right data structure.

Queries: Process query research is a method for systematically implementing queries and supplying process queries in the process store.

- This framework highlights integrated process questions, including architecture and observed behavior, based on search process structure and behaviour.

- To better help strategic decision making and provide a business intelligence forum, the process query approach addresses the above differences.

Visualization: Visualizations stood out as important feature for BI system user. Visualizations organize and present knowledge in a way that people with all levels of BI skill sets can understand and use.

Analytic techniques:

Regression analysis: A statistical approach used to find the relationships between two or more independent and dependent variables is the regression analysis in industry. One variable is independent and is calculated by its effect on the other dependent variables. There are many regression models, the most common of which are: Linear regression model, Nonlinear regression and time series regression (Anurag, 2020).

Machine learning: Machine learning is a type of artificial intelligence in which, without being specifically programmed to do so, a computer can perform tasks. Training a machine learning algorithm usually needs a lot of data that is purposefully "cleaned," or ordered and organized. The computer constructs an algorithm based on the patterns it recognizes by analyzing the data (and its labels, if applicable). It is then possible to apply the algorithm to completely different data sets (Pete Reilly, 2018).

# BI tools:

### ****Tableau****

Tableau is a fast, simple, and intuitive tool for performing analytical operations for everyone.

For business departments, using data to drive business operations requires an environment that can be analyzed and processed graphically.

However, in many cases, it is necessary to have programming skills and knowledge, so there are many requirements for analysis and reports sent to the IT department or the specialized department.

Such requests, even if they are sent early, cannot perform the analysis immediately as they are related to manpower and time, which can lead to delays in operations ( decide) business.

On the other hand, if only reporting in excel is based on letters and numbers, it will be difficult to give a visual view.

Currently, there are many data analysis tools called BI (Business Intelligence) solutions. In which Tableau is a BI tool used by many people inside and outside Japan.



Figure 1: tableau

**ADVANTAGES:**

* Free version available.
* Can interact with any type of data from Excel, Data Warehouse to Website Data.
* Real-time data updates.
* Make Data Visible in many ways like Charts or even a Dashboard - Better than any other software on the market.
* Tableau's Big Data processing system is very powerful.

**Defect:**

The Source of the Data must be processed in order to use Tableau's full power.

(A1 digihub, 2019)

### ****Python****

Python is a popular and very interesting scripting language. It is not a language with fast execution speed like Assembly, C, C ++ ...

It was chosen as the first programming language to teach novices or teenagers. Google, Microsoft, and many corporations and companies use IT to operate their service systems.

Course researchers, big data analysts also like to use Python for their work, because it works, doesn't get fancy, learns quickly, and does it.

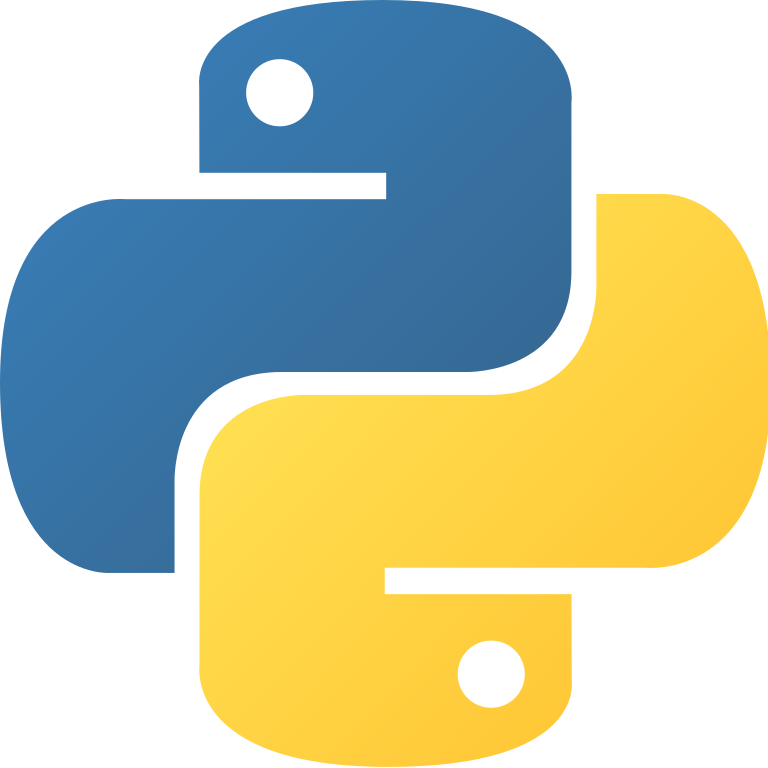


Figure 2: Python

ADVANTAGES

* Rated to be easy to learn, easy to write, easy to maintain, and offered as Open Source (Free).
* There are good machine learning libraries: Scikitlearn, Theano, Tensorflow, and Keras.
* Ability to collect on many platforms such as SQL server, MongoDB, JSON files.
* Handles Text Data very well.

(A1 digihub, 2019)

### ****Excel****

Excel is in the Microsoft Office suite of office tools, including many softwares that support writing, presenting, managing emails or spreadsheets like Excel.

This software helps to create spreadsheets, with features and tools to help users calculate data quickly and accurately with the number of millions of cells.



Figure 3: Excel

Important point:

* Is the most basic and popular tool for Data Analysis.
* Used in data analysis in all professions.
* There are advanced analysis features that help model data such as automated relational creation, DAX solution, and time grouping.
* Whether you're a Sas, R or Tableau expert, you still need Excel.

(A1 digihub, 2019)

# Explain dataset

The data file is presented in excel, with full information from many famous soccer players in the world. Data is actually taken, including 89 columns Age, Nationality, Overall, Potential, Club, Value, Wage, Preferred Foot, International Reputation, Weak Foot, Skill Moves, Work Rate, Position, Jersey Number, Joined, Loaned From, Contract Valid Until, Height, Weight, LS, ST, RS, LW, LF, CF, RF, RW, LAM, CAM, RAM, LM, LCM, CM, RCM, RM, LWB, LDM, CDM, RDM, RWB, LB, LCB, CB, RCB, RB, Crossing, Finishing, Heading, Accuracy, ShortPassing, Volleys, Dribbling, Curve, FKAccuracy, LongPassing, BallControl, Acceleration, SprintSpeed, Agility, Reactions, Balance, ShotPower, Jumping, Stamina, Strength, LongShots, Aggression, Interceptions, Positioning, Vision, Penalties, Composure, Marking,

StandingTackle, SlidingTackle, GKDiving, GKHandling, GKKicking, GKPositioning, GKReflexes, and Release Clause. And the famous 18208 soccer player. However, the used data is not clean, so our team needed to process this pile of data before using it for the Design dashboards.

# Explain pre-process steps on dataset

Before designing the dashboards we need to clean up this pile of data. At the request of the instructor, we need to remove the columns after the ID, Flag, club logo. In the value column, convert this quantity into a quantity that can be calculated, leaving the unit. At the work rate column split into two separate columns. Placement, from the abbreviations, we will make a rewrite of the name according to its true meaning. Weight will be removed from units, height of the players will be converted to cm. In the player's index position, there are still numbers like "90 + 2" these numbers will be added up. Contract release will remove the unit part.

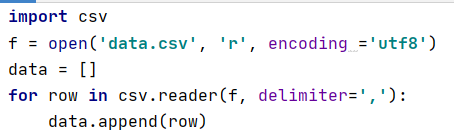
The first thing we download data is available to the player to the computer, use the code to open the data:

Figure 4: code open

Delete columns:

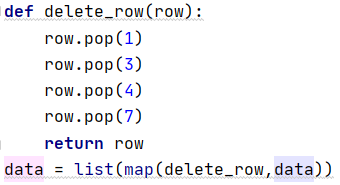


Figure 5: Code delete columns

Change symbols to letters with the same meaning:

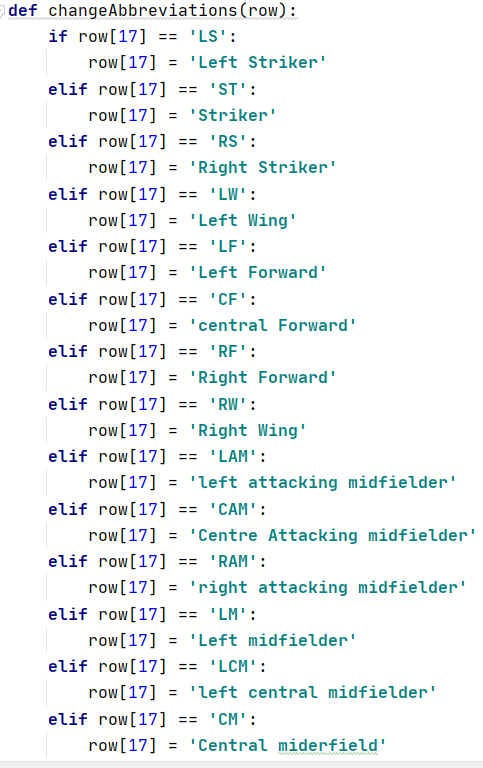


Figure 6: code Change symbols to letters with the same meaning

Correct value:

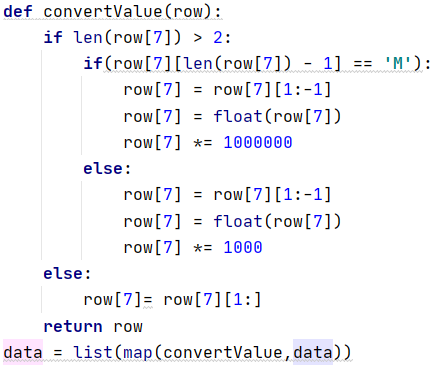


Figure 7: code Correct value

Reputation of wage:

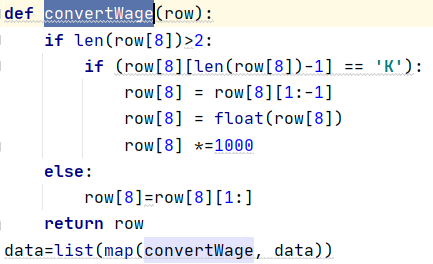


Figure 8: code Reputation of wage

Edit Height:

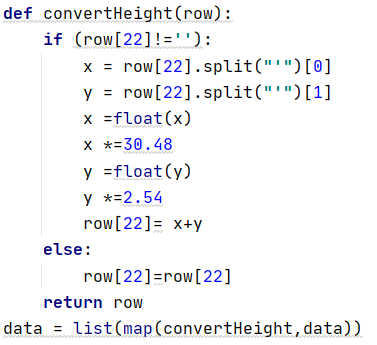


Figure 9: code Edit Height

Fix weight:

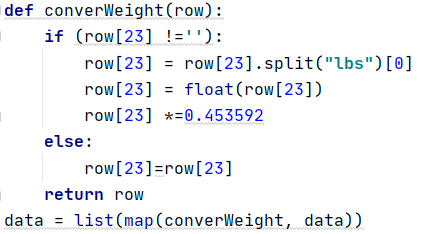


Figure 10: code Fix weight

Indicates the position of each player on the field:

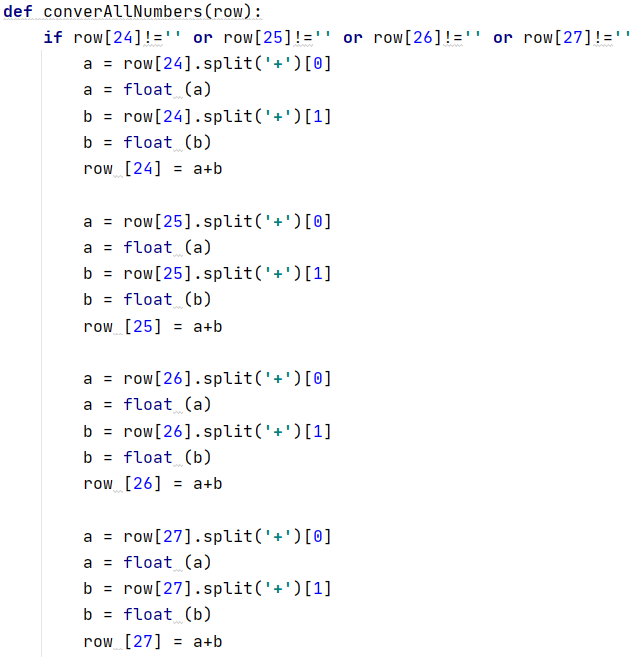


Figure 11: code Indicates the position of each player on the field

fix Release Clause:

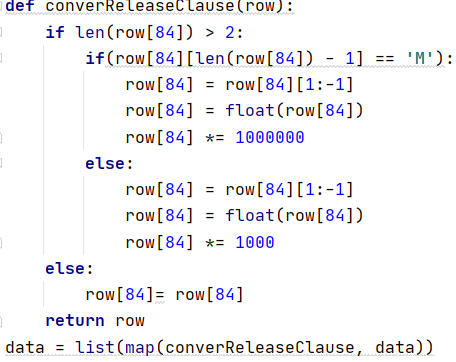


Figure 12: code fix Release Clause

Split the column in two:

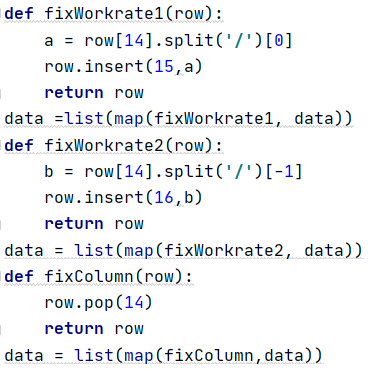


Figure 13: code Split the column in two

Export the cleaned data to a new excel file:

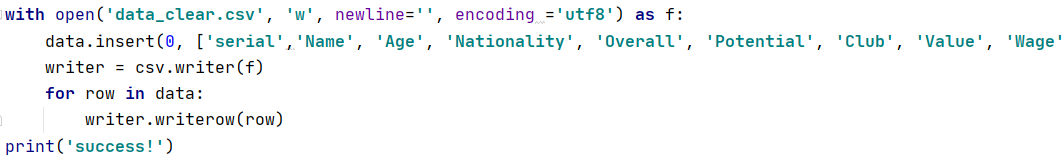


Figure 14: code Export the cleaned data to a new excel file

# Design Dashboards:

In this report, I made 2 dashboards on 2 different topics, I applied previously prepared data to create these 2 dashboards.

Dashboard 1:

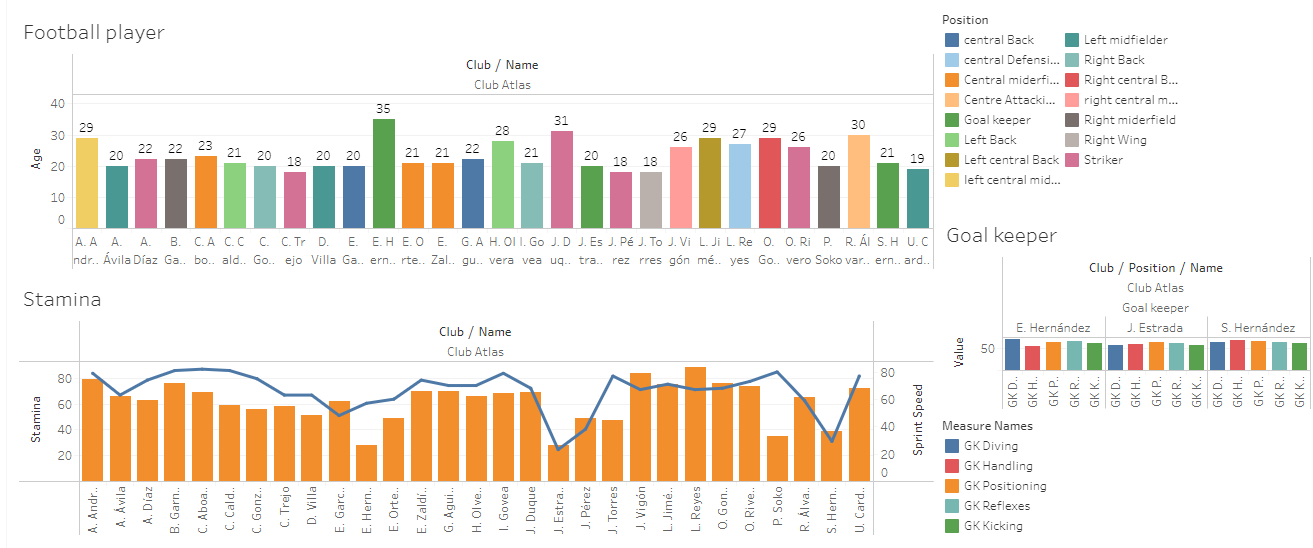


Figure 15: Dashboard 1

When my superiors assigned us to find the central midfielder, I created this dashboard to help my superiors make the right decisions. In this dashboard includes information related to money such as player value, salary, contract release value ... With this Dashboard, superiors can find suitable players that are valuable to the team.

**Contain charts:**

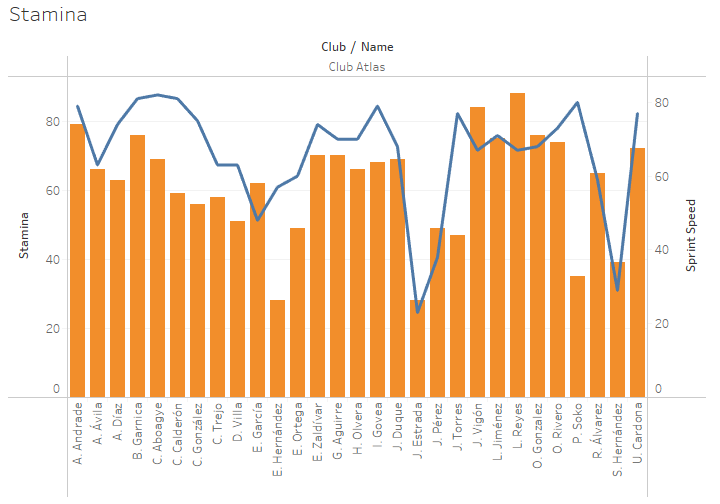


Figure 16: Aim to find out which players have low stamina and speed on the team

Aim to find out which players have low stamina and speed on the team

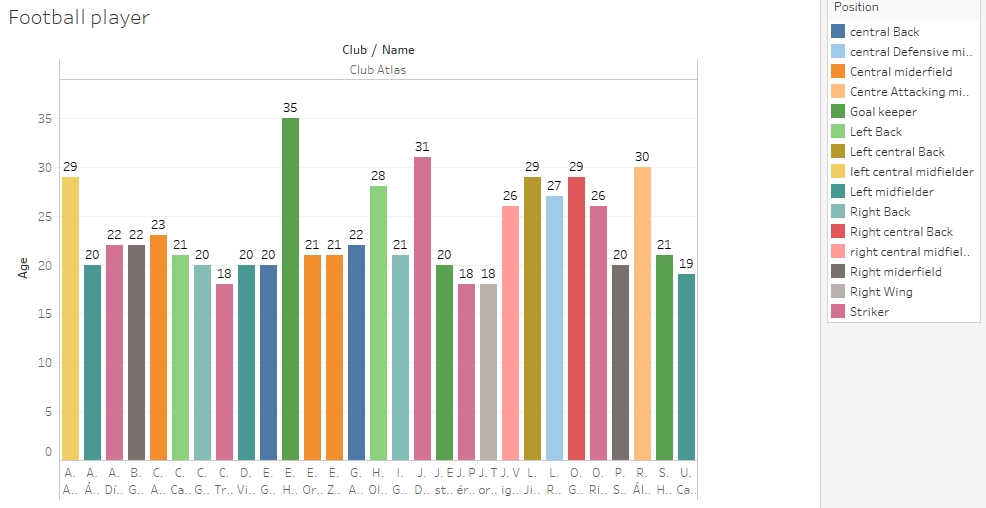


Figure 17:After I discovered those players, I looked at where they played, and what age they were

After I discovered those players, I looked at where they played, and what age they were

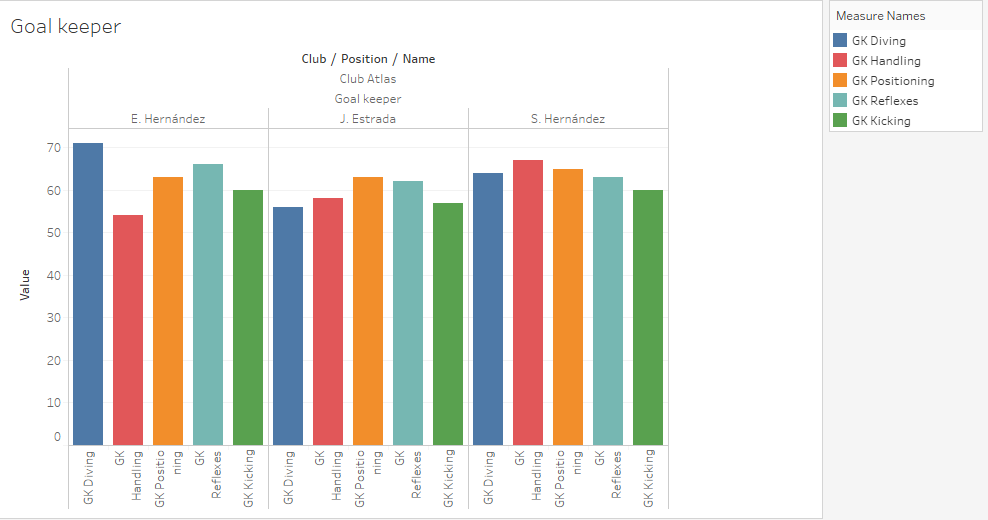


Figure 18: After I have identified the players with low stats, I find out if they are still performing well in the stats

After I have identified the players with low stats, I find out if they are still performing well in the stats.

**Interaction on dashboard:**

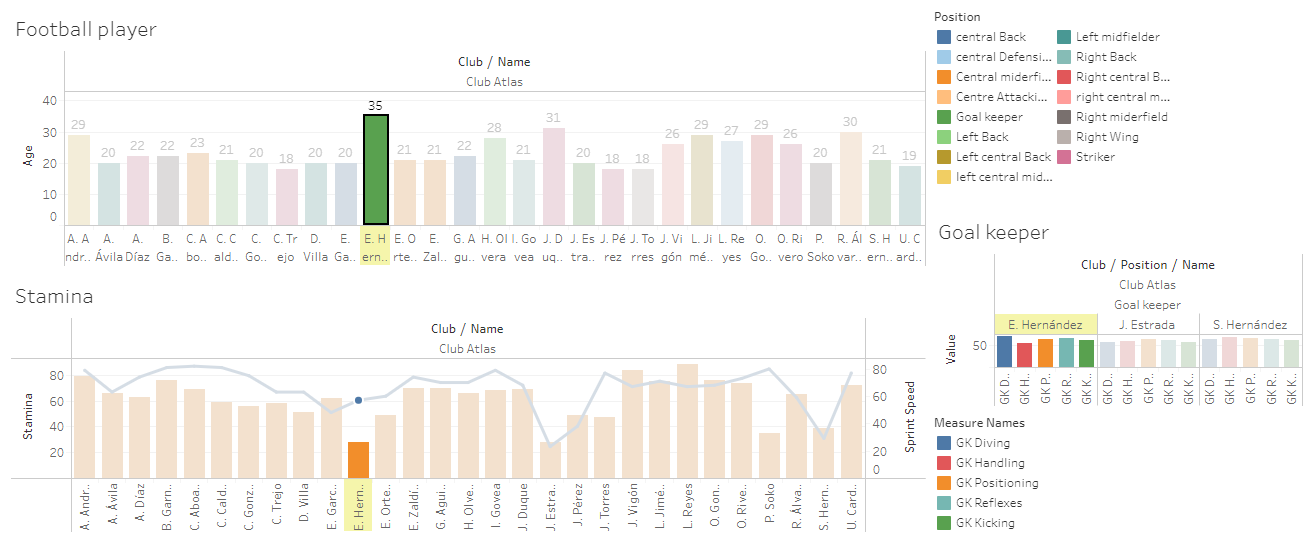


Figure 19: Interaction on dashboard

**Result:**

Those with low stamina and speed are in the goalkeeper position. That is not an indicator too important for the goalkeeper. There is a goalkeeper who is 30 years old but still has better stats than younger goalkeepers. In addition, there is a right-wing midfielder who has low physical strength, but the speed is not inferior to his teammates. With this result, superiors can completely make decisions to reorganize the team

**Conclusion:**

The goalkeeper is still good, but he is old, it is not certain that he will play well in the future, the superiors will consider and decide whether to keep the player or not. In addition, young players should also have physical tranning sessions or also sell or buy better players.

As for the right winger, he is physically weak but fast, can be physically fit and put on the pitch for the last 20 minutes

**Suggestions:**

* Older players can play until the end of their contract or sell if necessary
* Young goalkeepers also have pretty good stats, can keep and train their fitness or sell to buy other goalkeepers
* A winger with a low stamina score could put in the 80th minute to make a spike. He is very young and can continue to train

Dashboard 2:

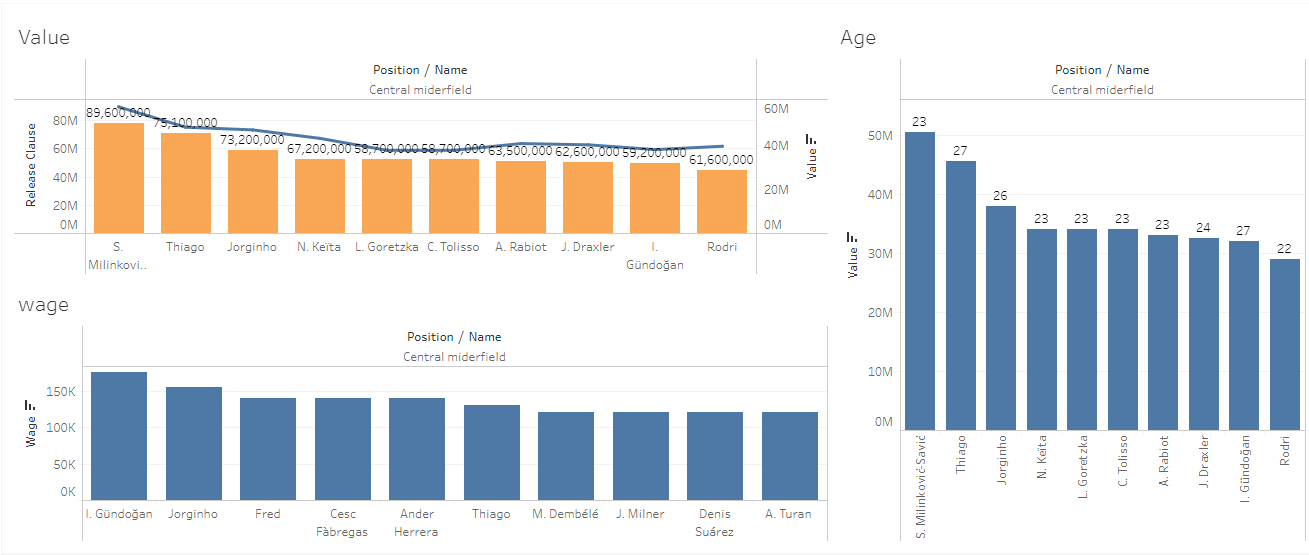


Figure 20: Dashboard 2

Suppose the coaching staff wants their team to have similar skills among their players. He wants to transfer a few players in the squad based on criteria: especially age, fitness and speed. I created a dashboard to analyze the players and give my superiors options for them to consider whether to keep holding those players.

**Contain charts:**

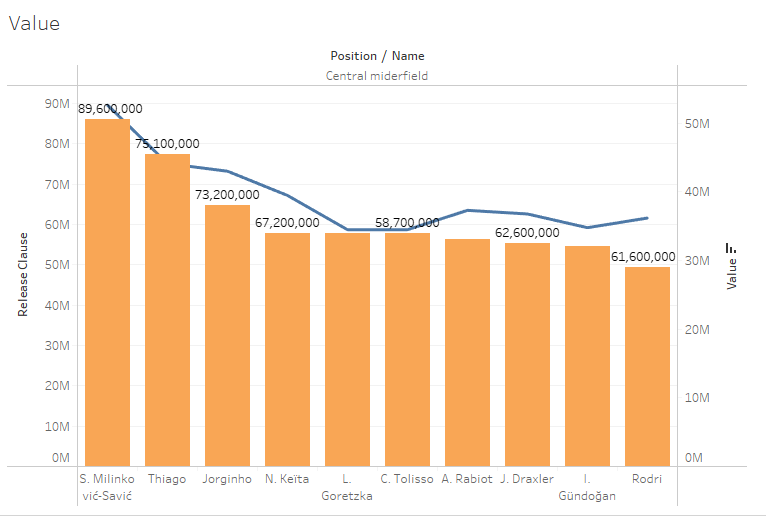


Figure 21: Contain charts

This chart is used to find the top midfielders according to the player's value criteria, and also displays the contract release price of that player.

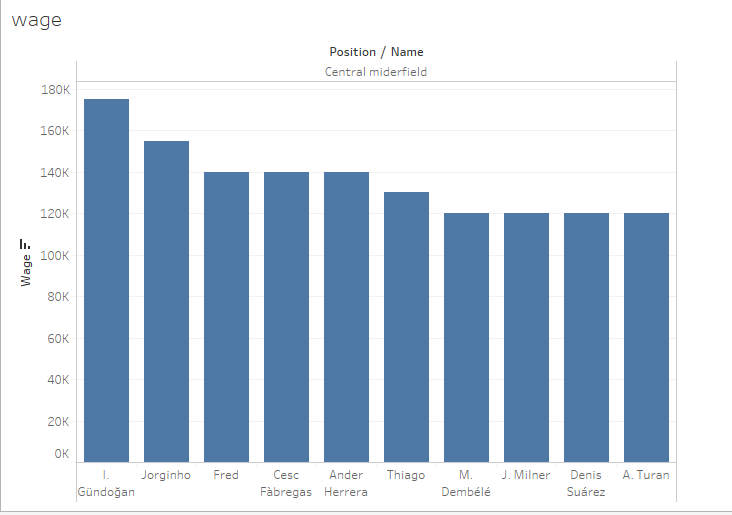


Figure 22: wage

This chart is used to show the salaries of the players with the highest value

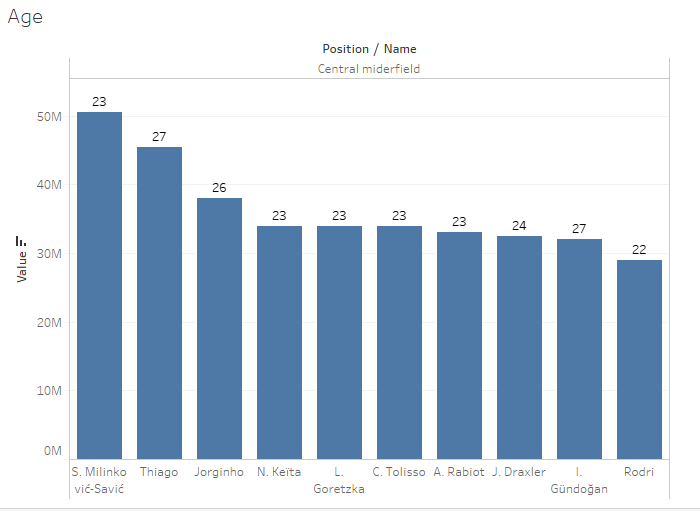


Figure 23: Age

This chart is used to show the age of the players with the highest value, whether with their value when buying, how long they can contribute to the team.

**Interaction on dashboard:**

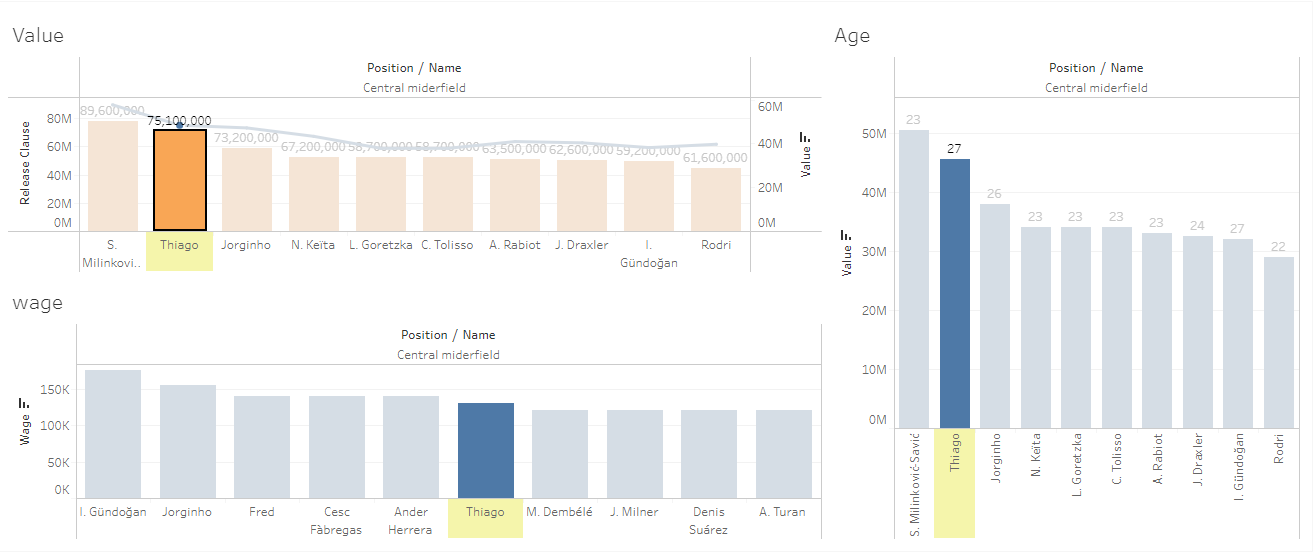


Figure 24: Interaction on dashboard

**Result:**

It can be seen that there is a number 1 player with both high value and low salary. Moreover, he is very young, and can completely dedicate himself to the team. However, he does have a sizable amount of contract release. The remaining players have less money released, but there are a few salaries in the top. Thanks to this result, the superior will consider with the money, which player is most valuable for the team

**Conclusion:**

For the first player, having a player of exceptionally high value and a non-top salary is a good option. However, the value of releasing the contract is quite high, the selection of players is not only in the player's value but also in the player's skill. Therefore, it is necessary to further analyze the indicators of that player's skills.

Besides, there is also 1 player with high value and the amount of contract release is also less, the superior can fully consider these two cases.

**Suggestions:**

# It is possible to choose the first player, because his value is absolutely superior to the rest, his salary is not in the top. However, if the economy is not enough, the team can choose other players with much less money

# Discuss the legal issues involved in exploiting user data for business intelligence

When using user data for market intelligence, there is still concern about data protection and infrastructure reliability.(Beleuta, 2017)

The European Union has established a law that limits the storage and processing of personal data, called the General Data Protection Regulation or GDPR after a number of scandals about voluntarily collecting user data took place in 2018.

The Law focuses on ensuring that consumers identify, understand and agree to the data they receive. There's a fine breach of "up to EUR 10 million."

User data collection must also conform with regulatory provisions in order not to be embroiled in legal problems that may inflict financial harm, decrease the overall integrity of the company and weaken its image.

The measures that can be taken to comply to legal and ethical data collections constrains are, but not limited to:

o Be upfront and transparent regarding data collection to customers and users

o Apply security measures to safeguard collected user data

o Do not use collected information for any other purpose than the intended and

informed to the customer.

o Any changes to data collection handling policy should be informed to user.

# Discuss how business intelligence tools can contribute to effective decision-making

Data visualization is very important because people react to images much better than other forms of presentation such as symbols, letters, numbers…. So, working with Data visualization tools like Tableau will help people understand data better, as it allows one to access the amount of data with easy-to-understand graphical images. This is the simplest and most powerful way to present any data. From there, it helps to make the most intuitive analysis and synthesize all those analyzes and make the most effective decisions.

An example:

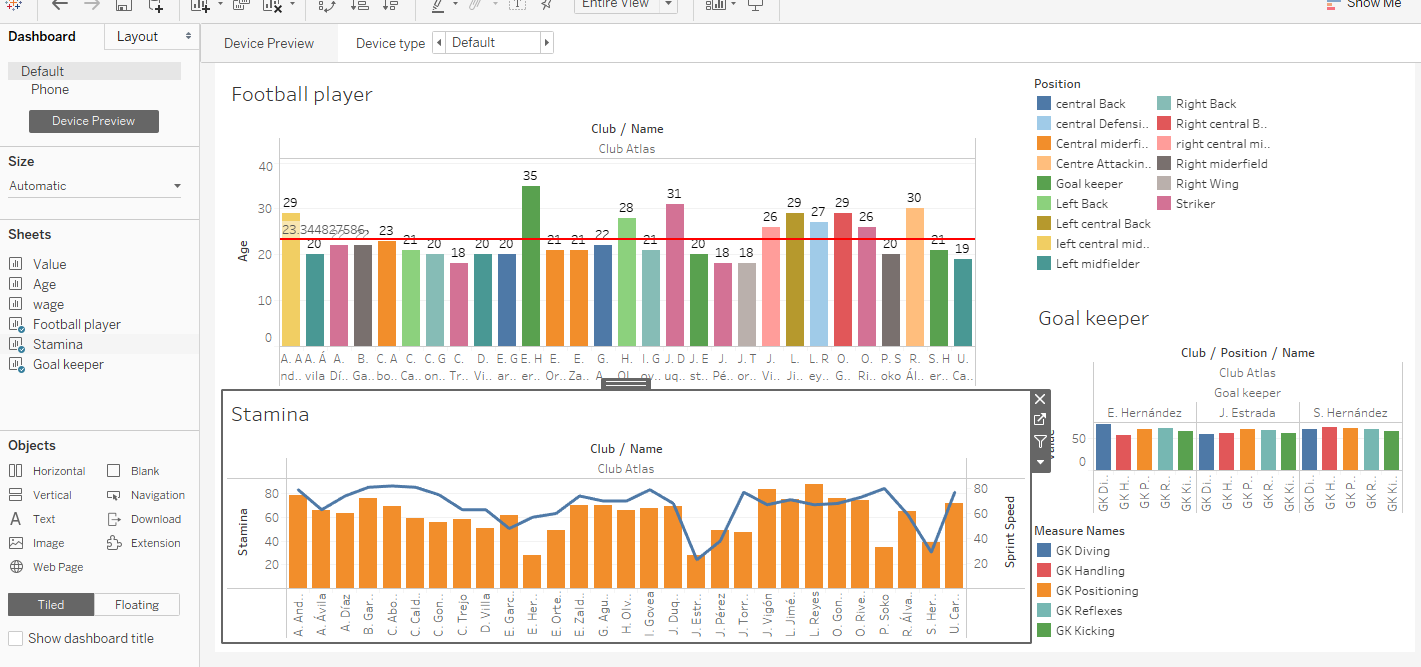


Figure 25: Dashboard

To see if a player is still capable of contributing to the team and make a decision whether to extend his contract or not. Our team used tableau to create an upper dashboard to show data such as name, age, fitness and running speed, his specific location and his chi stats. details.

Looking at Stamina chart we can easily see that 4 people with weak strength are E.Hernandez (28), J. Estrada (28), P. Soko (35) and S.Hernandez (39).

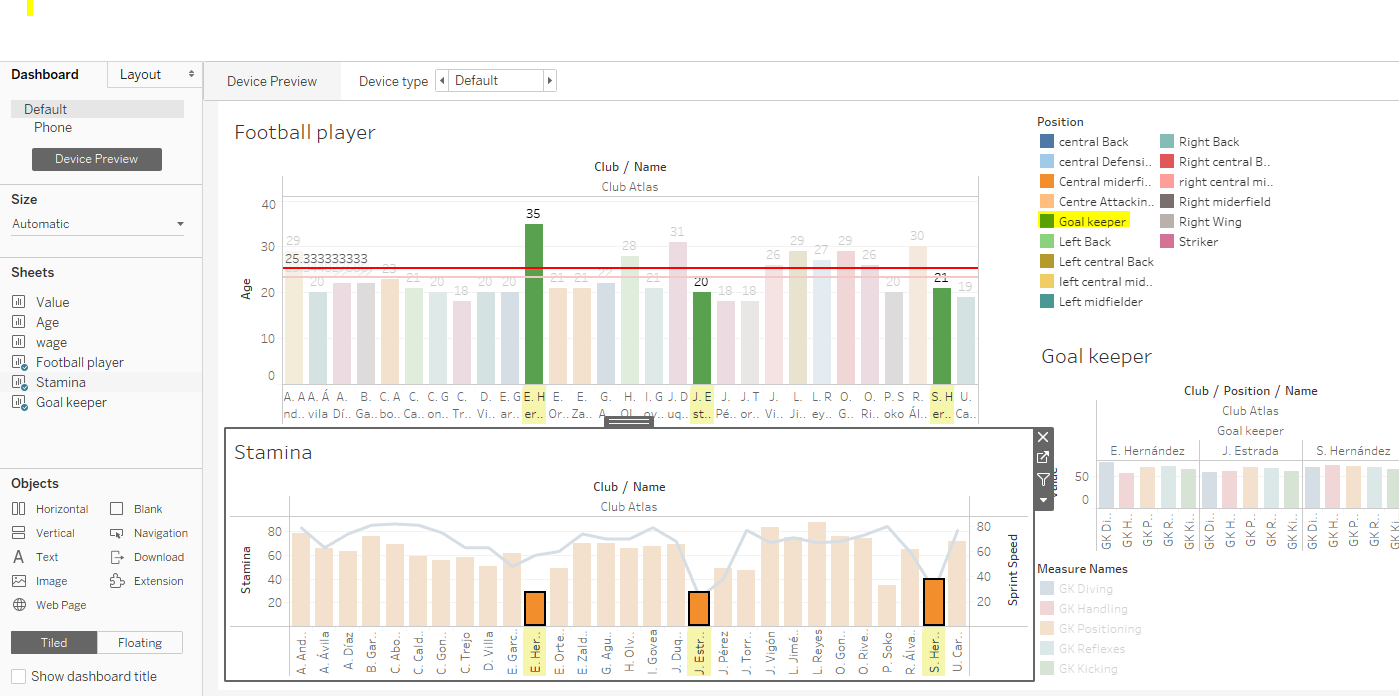


Figure 26: Highlight data

However, we can see that 3 out of 4 are goalkeepers (GK), so low stamina is not a problem, but there is another problem that is E.Hernandez is 35 years old, but 2 keepers The other subject is at a younger age than J. Estrada 20 years old and S.Hernandez 21 years old, looking at the chart Goal keeper, they are all equally good goalkeepers with nearly the same index. Thus, based on this dashboard, it is possible to make a priority decision to sign contracts with two potential young goalkeepers, J. Estrada and S.Hernandez.

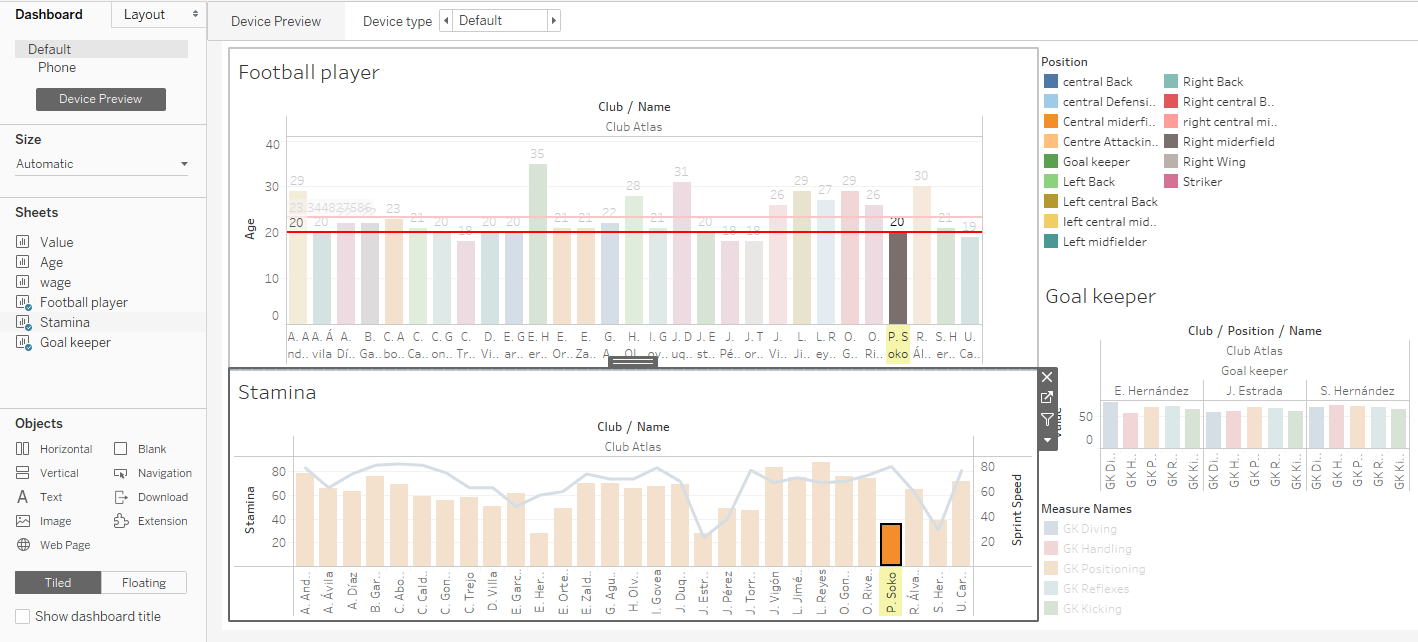


Figure 27: Highlight Data

The other person on the list of weak fitness is P. Soko (35) and he is a midfielder, a position in need of strength, so the dashboard helps to see the problem, come here can take made a decision to dismiss this player. However, when looking at the green line of speed (in the Stamina chart), P. Soko has a very high acceleration index, which is very useful. So the problem here is only this player's fitness, based on these things we can make a more accurate decision that is not to fire this player, but instead will focus on trainning to improve his body. force for this player.

# Feedback/comment

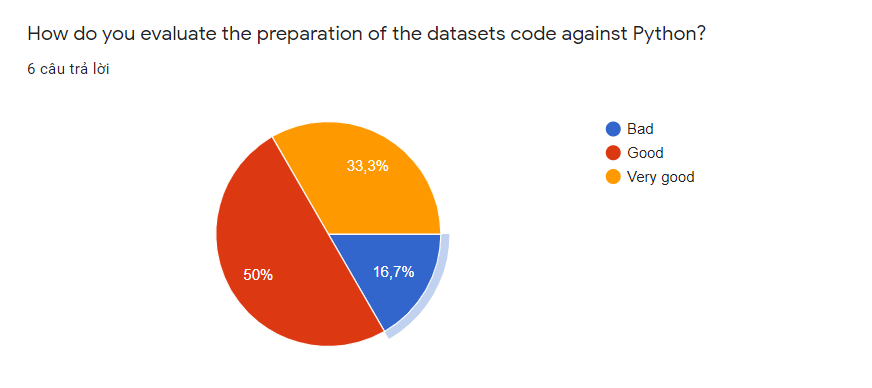


Figure : datasets code against Python

Most people are satisfied with the code. 50% percent shows that the code runs well. 33.3% failed to run code very well, but 16.7% found that it was not satisfied. The reason may be that the group's code is still simple.

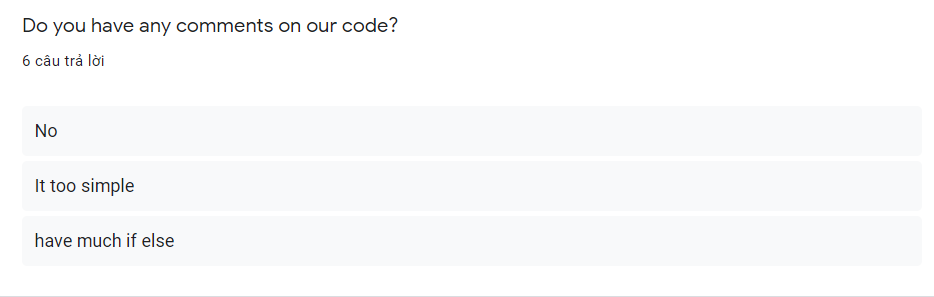


Figure : comments on our code

Most people find simple code has too much if else. The team will overcome in the near future.

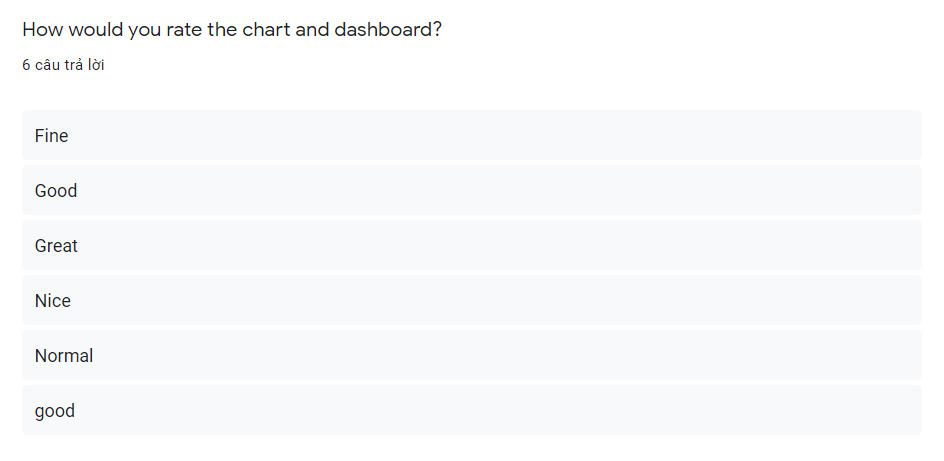


Figure : rate chart and dashboard

Everyone sees good charts and dashboards.

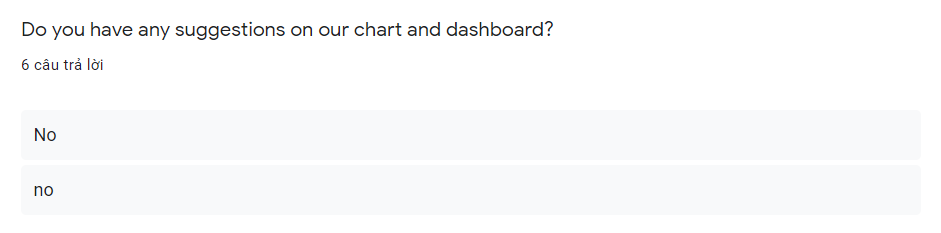


Figure : suggestions on our chart and dashboard

Most people do not have any opinion on dashboard.

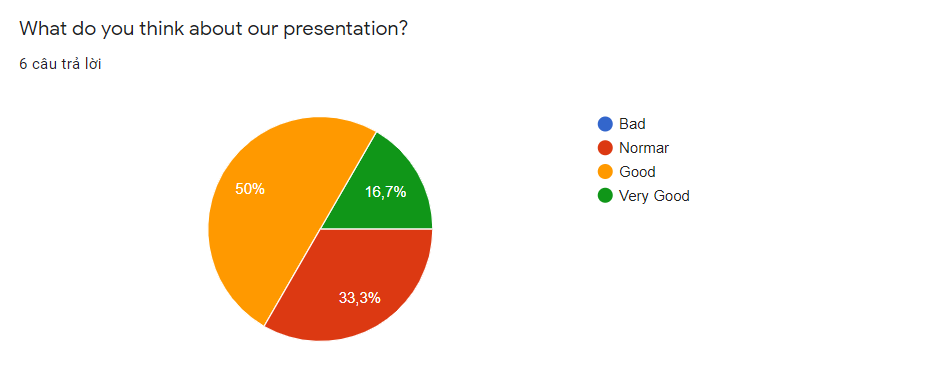


Figure :our presentation

33.3% of people gave the opinion of the normar group presentation, the reason could be that the group had been presenting for too long without outstanding emphasis. Will overcome in the near future.

Summary :

Most people are satisfied with the group's work. However, the group still has some shortcomings that need to be overcome as follows:

- Code needs to be less if else and optimize.

- Need to overcome more about presentation skills.

# Bibliography

*A1 digihub*. (2019, 6 11). Retrieved from Phân tích dữ liệu? Top 10 công cụ tốt nhất cho Nhà Phân tích Dữ liệu: https://a1digihub.com/top-10-cong-cu-tot-nhat-cho-nha-phan-tich-du-lieu/?fbclid=IwAR2eXyTEcnjvh4b0mFtHU9OVxVVYkjzyqFC46VeLie9E-uRIRyJ1RHOtYE8

*GURU99*. (n.d.). Retrieved from What is Business Intelligence? Definition & Example: https://www.guru99.com/business-intelligence-definition-example.html#4

Beleuta, V., 2017. Data privacy and security in Business Intelligence and Analytics, Barcelona: Universitat

Polit`ecnica de Catalunya.

https://en.wikipedia.org/wiki/General\_Data\_Protection\_Regulation