

VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY- UNIVERSITY OF
SCIENCE
FACULTY OF INFORMATION TECHNOLOGY



April 7, 2025

Data Visualization

Report lab 03

Lecturers and Teaching assistants:

1. Mr. Bui Tien Len
2. Mr. Vo Nhat Tan

Students:

1. Quach Tran Quan Vinh - 22127460
2. Nguyen Hoang Trung Kien - 22127478

Contents

1	Project and team progress	2
1.1	Overall progress	2
1.2	Team progress	2
2	Exploring Power BI	3
2.1	Introduction Power BI [1]	3
2.2	Functions and example usage [2]	3
2.2.1	Import data	3
2.2.2	Transform data	5
2.2.3	Write expressions in DAX	6
2.2.4	Data modeling	8
2.2.5	Data visualization	9
2.2.6	Dashboard	10
2.2.7	Responsive devices	11
2.2.8	Auto-generated narrative	12
3	Main visualization task	13
3.1	Data collection	13
3.2	Data preprocessing	13
3.3	Data exploration	13
3.4	Data visualization	15
3.4.1	Implementation Overview	15
3.4.2	Dashboards and Analysis Objectives	16

1 Project and team progress

1.1 Overall progress

#	Tasks	Progress
1	Data collection	100%
2	Data preprocessing	100%
3	Data dashboard visualization and analysis	100%
3	Discover Power BI	100%

1.2 Team progress

#	Student's ID	Name	Tasks	Progress
1	22127460	Quach Tran Quan Vinh	Data collection	100%
2	22127478	Nguyen Hoang Trung Kien	Data preprocessing	100%
3	22127478	Nguyen Hoang Trung Kien	Data exploration	100%
4	22127478	Nguyen Hoang Trung Kien	Write introduction to power BI	100%
5	22127460	Quach Tran Quan Vinh	Write functions and example usage of Power BI	100%
6	22127478	Nguyen Hoang Trung Kien	Design overview dashboard, dashboard 1 and 2	100%
7	22127460	Quach Tran Quan Vinh	Design dashboard 2 and 3	100%
8	22127478	Nguyen Hoang Trung Kien	Make and answer questions 1, 2, 3	100%
9	22127460	Quach Tran Quan Vinh	Make and answer questions 4, 5, 6	100%
10	22127460 and 22127478	Quach Tran Quan Vinh and Nguyen Hoang Trung Kien	Write report	100%

2 Exploring Power BI

2.1 Introduction Power BI [1]

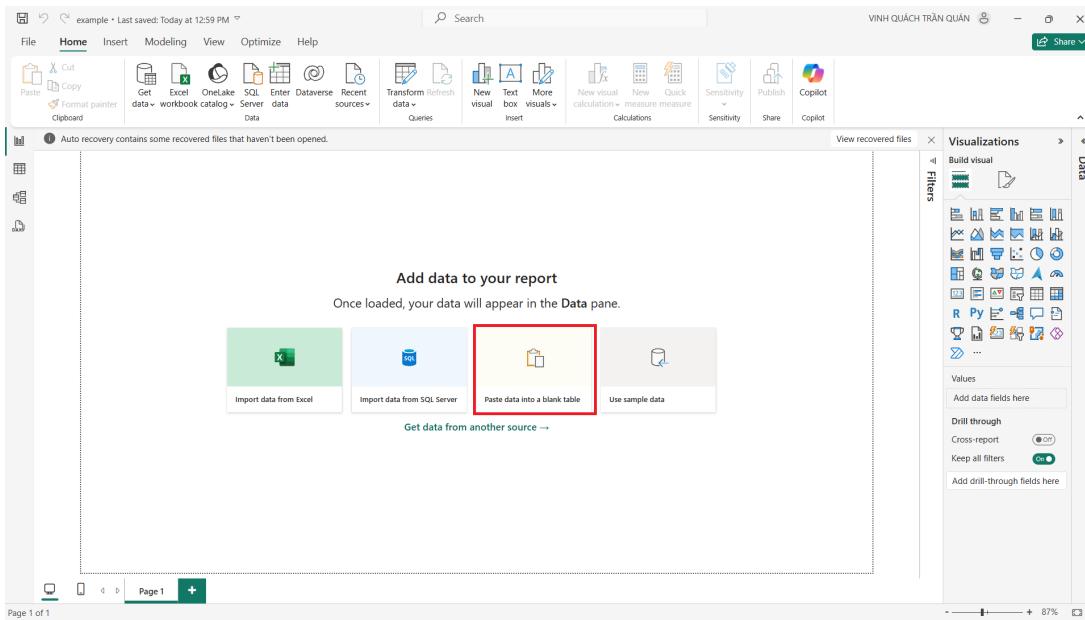
Power BI is a powerful platform that enables users to perform various data-related tasks, including analysis, visualization, and reporting. It provides a wide range of functions to create structured and visually appealing charts, graphs, and dashboards. With its friendly interface and drag-and-drop functionality, Power BI makes it easy for users to transform raw data into meaningful insights. Additionally, it supports data modeling, custom calculations using DAX (Data Analysis Expressions), and integration with multiple data sources, allowing users or even businesses to make informed decisions based on real-time data.

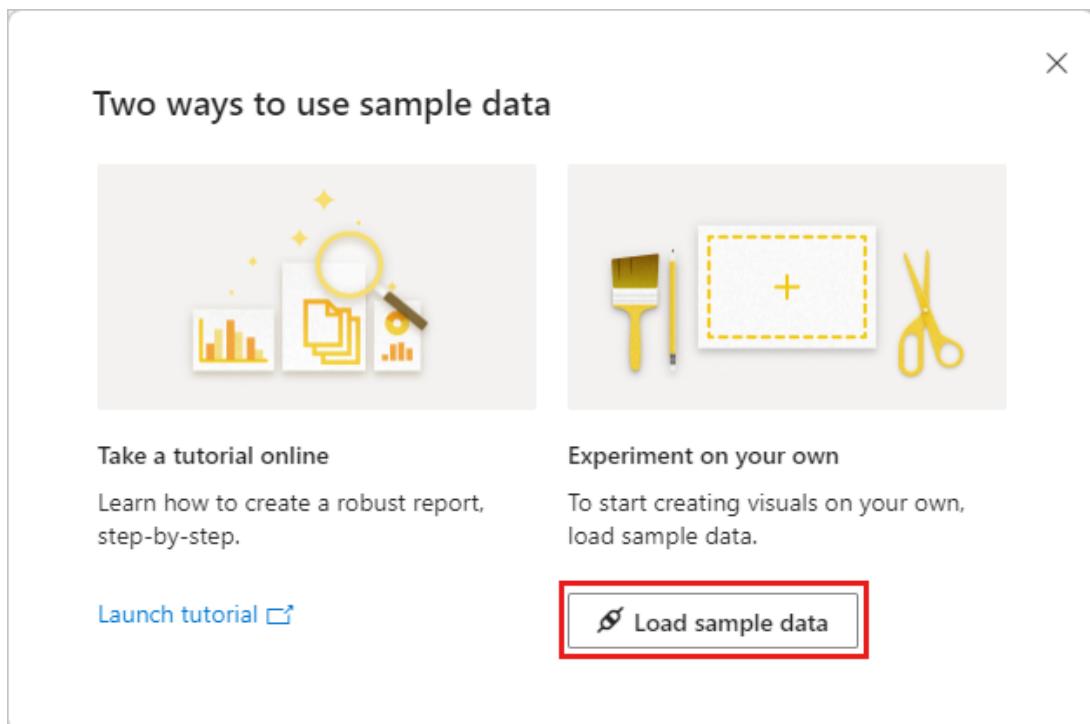
2.2 Functions and example usage [2]

Power BI provides multiple features to visualize data. In this section, we will use Financial Sample Excel Workbook dataset which is available in Power BI.

2.2.1 Import data

In this subsection we will import the sample data - Financial Sample dataset. To import the existing sample data, we select





The Navigator dialog box shows a file list on the left with 'Financial Sample.xlsx [2]' and a table preview on the right titled 'financials'. The 'financials' sheet is selected, indicated by a red box around its row in the list. The table preview shows data for various segments, countries, products, and discount bands. At the bottom are 'Load', 'Transform Data', and 'Cancel' buttons.

Navigator

Display Options

Financial Sample.xlsx [2]

financials

Sheet1

financials

Segment	Country	Product	Discount Band	Un
Government	Canada	Carretera	None	
Government	Germany	Carretera	None	
Midmarket	France	Carretera	None	
Midmarket	Germany	Carretera	None	
Midmarket	Mexico	Carretera	None	
Government	Germany	Carretera	None	
Midmarket	Germany	Montana	None	
Channel Partners	Canada	Montana	None	
Government	France	Montana	None	
Channel Partners	Germany	Montana	None	
Midmarket	Mexico	Montana	None	
Enterprise	Canada	Montana	None	
Small Business	Mexico	Montana	None	
Government	Germany	Montana	None	
Enterprise	Canada	Montana	None	
Midmarket	United States of America	Montana	None	
Government	Canada	Paseo	None	
Midmarket	Mexico	Paseo	None	
Channel Partners	Canada	Paseo	None	
Government	Germany	Paseo	None	
Channel Partners	Germany	Paseo	None	
Government	Mexico	Paseo	None	
Midmarket	France	Paseo	None	

Load **Transform Data** **Cancel**

2.2.2 Transform data

- We can modify the data such as (preprocessing; handling missing data; remove columns, rows; etc.) in **Transform** tab.

The screenshot shows the Power BI desktop interface. The 'Transform' tab is selected in the ribbon. The main workspace is empty, showing a placeholder 'Build visuals with your data'. To the right, the 'Data' pane lists various data fields categorized under 'Values' and 'Drill-through fields here'. A red box highlights the 'Transform data' icon in the ribbon.

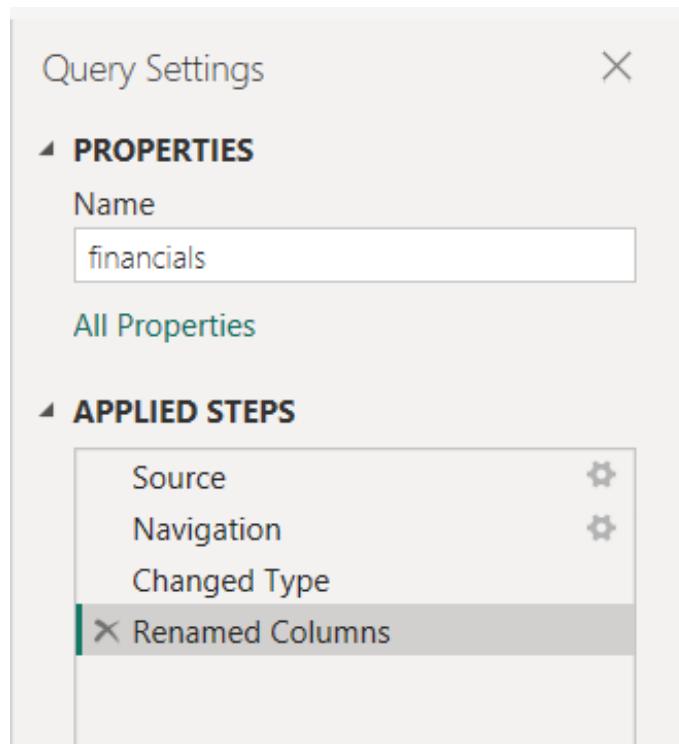
- The units sold are in decimal form. It doesn't make sense to have 0.2 or 0.5 of a unit sold, so we changed it to a whole number.

The screenshot shows the Power BI desktop interface with the 'Transform' tab selected. The main workspace displays a table with columns: 'Discount Band', 'Units Sold', 'Manufacturing Price', 'Sale Price', and 'Profit'. The 'Properties' pane on the right shows 'Changed Type' applied to the 'Units Sold' column. A red box highlights the 'Transform' tab in the ribbon.

- Let's shorten the column name from **Month Name** to just **Month**. Double-click the **Month Name** column, and rename it to **Month**.

	A	B	C	Month Name	D	E	F
1	1	January			1	2	3
6	6	June					
6	6	June					
6	6	June					

4. You can see that each transformation has been added to the list under **Query Settings** in **APPLIED STEPS**.



5. Back on the **Home** tab, select **Close & Apply**. Our data is almost ready for building a report.

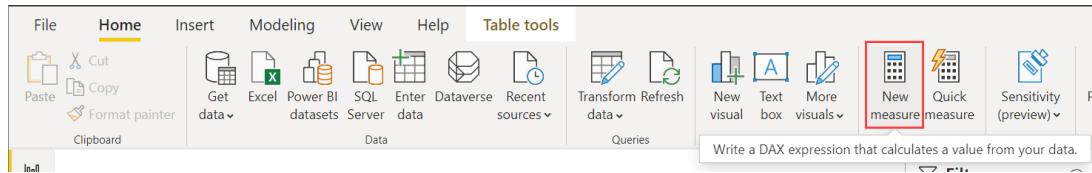
2.2.3 Write expressions in DAX

Writing measures and creating tables in the DAX formula language is super powerful for data modeling. There's lots to learn about DAX in the Power BI documentation. For now, let's write a basic expression to create a new measure, and another expression to create a whole table.

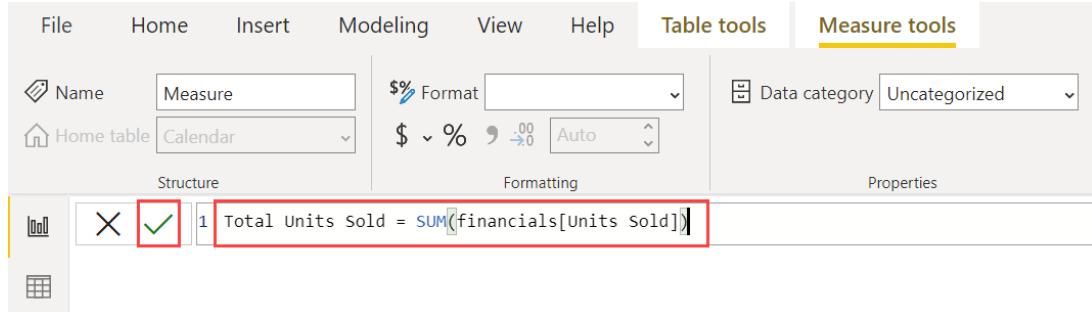
Create a new measure

We create a measure called **Total Unit Sold** which is sum of all unit sold values.

1. On the Home ribbon, select New measure.



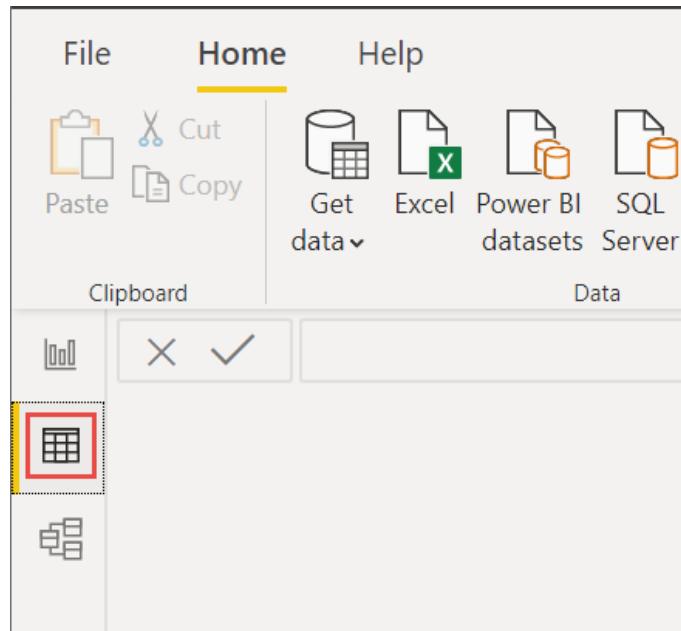
2. Write a DAX formula in the formula bar at the top of the report canvas to add all the numbers in the **Units Sold** column. Then select the check mark to commit.



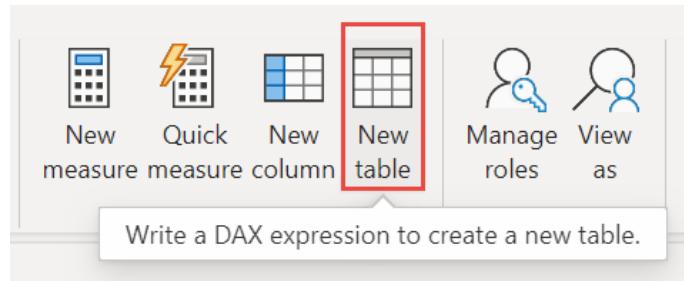
Create a new table

We create a table called **Calendar** to store date, so we are able to create a relationship between tables later on.

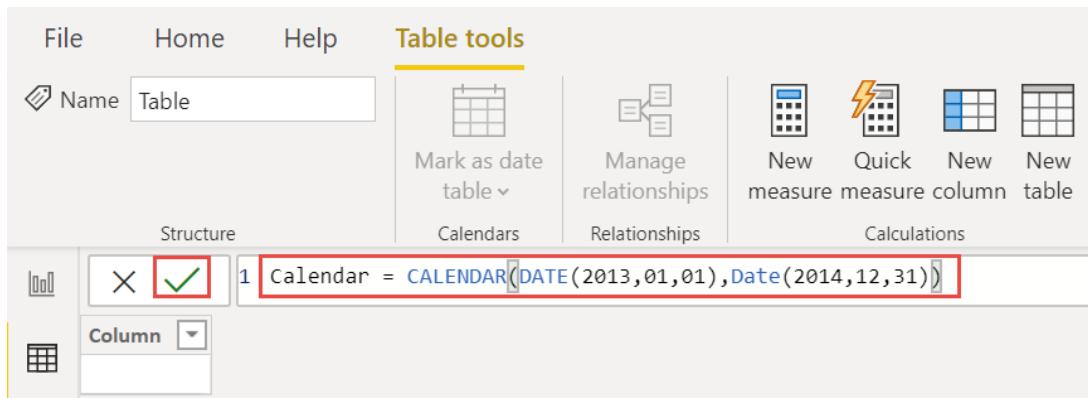
1. Now select the **Table** view icon on the left.



2. On the **Table tools** ribbon, select **New table**.

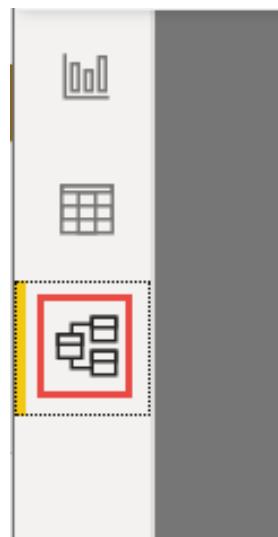


- As creating a measure, we enter the formula to the formula bar to create a table. We generate a **Calendar** table of all dates between January 1, 2013, and December 31, 2014.

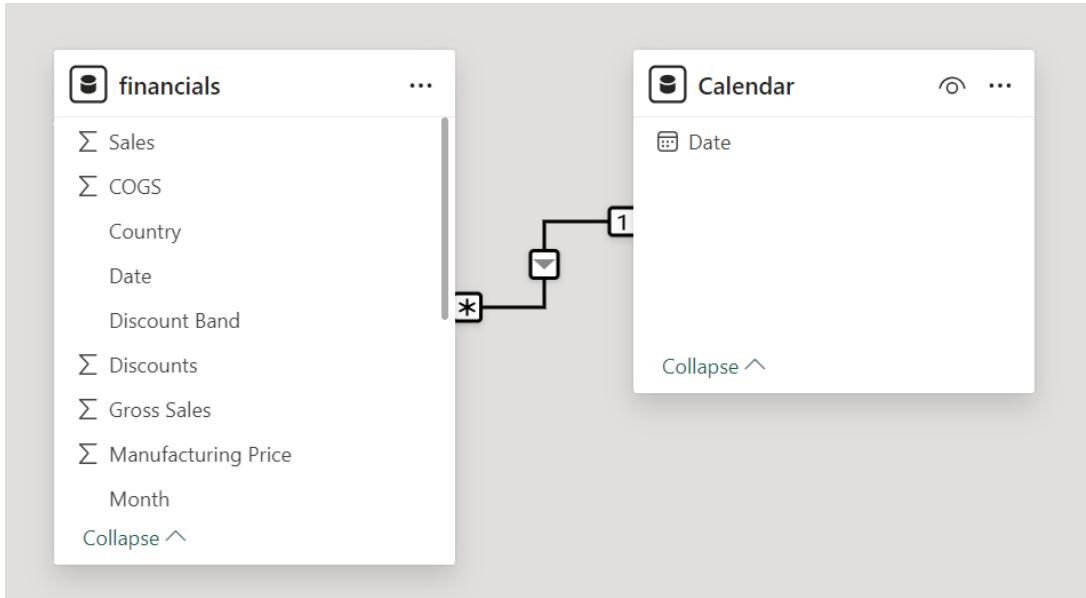


2.2.4 Data modeling

- First, we select **Model view** on the left.

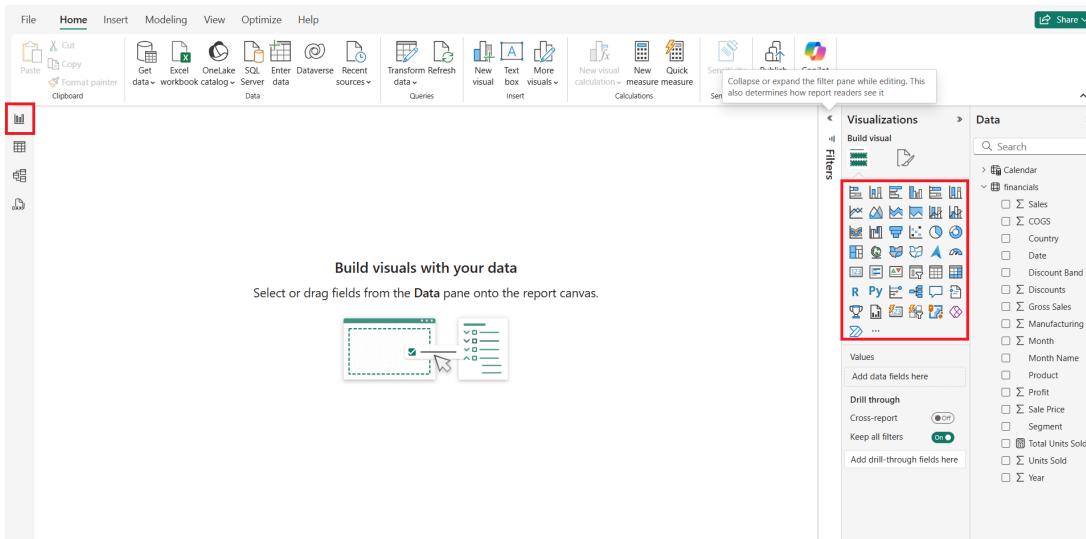


- Drag the **Date** field from the financials table to the **Date** field in the Calendar table to join the tables and create a *relationship* between them. In the **New relationship** dialog that opens, select **Save** to apply the selections that Power BI automatically filled in for you. As a result, we can see a *Many-to-one* relationship.



2.2.5 Data visualization

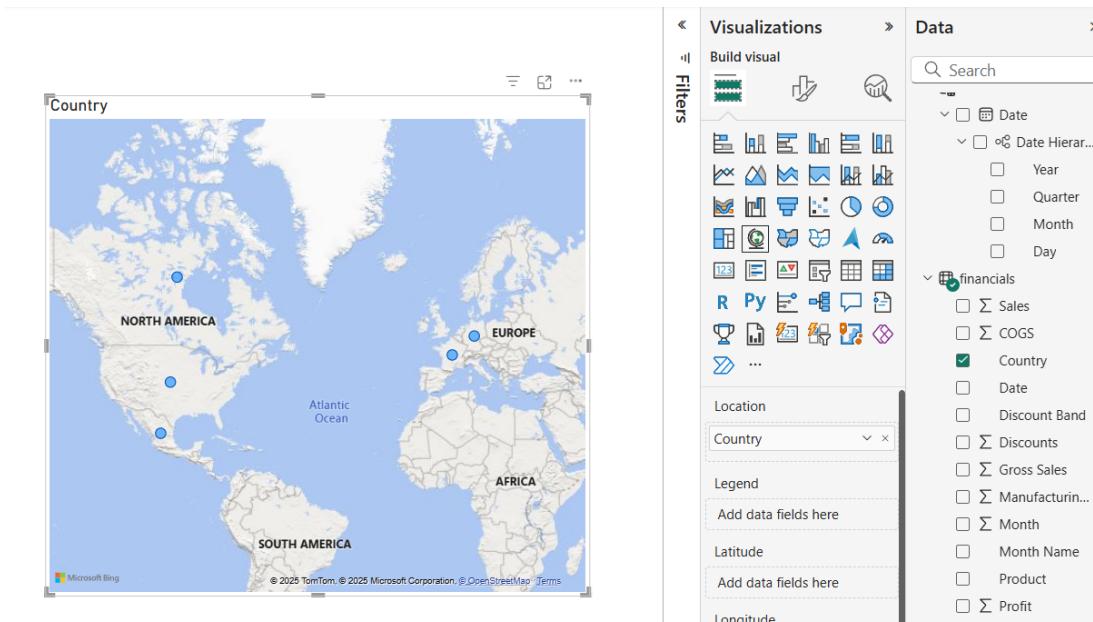
- Power BI offers multiple chart types by default. Users can access by selecting **Report View** on the left. Then, users can choose their preferred charts in the **Visualizations** pane on the right.



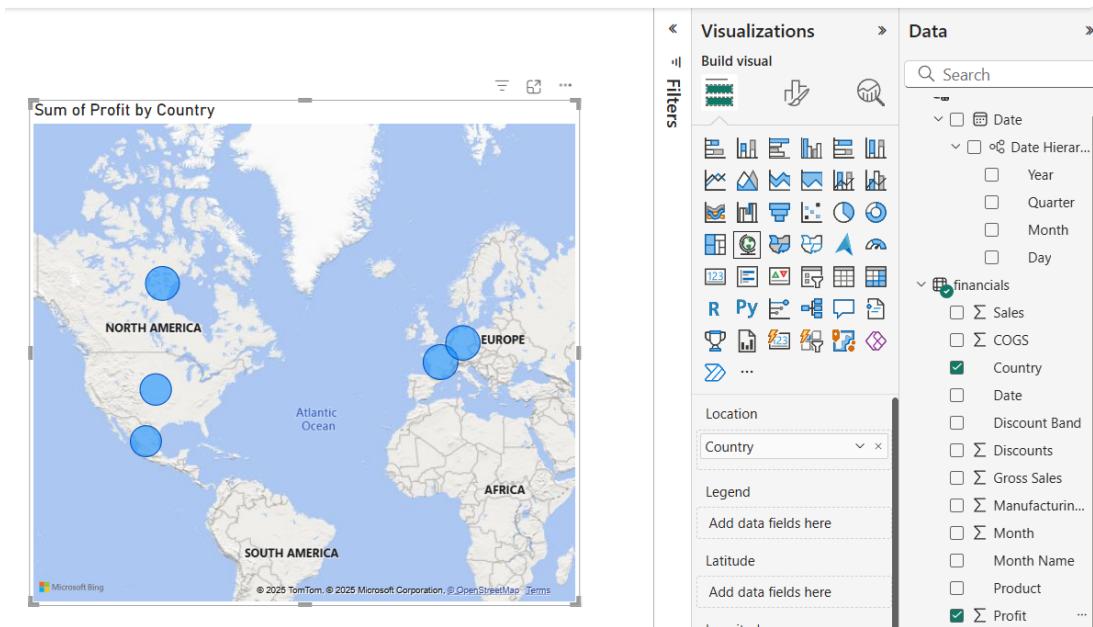
- Additionally, Power BI automatically generates appropriate charts based on different types of data.

For instance, we drag the **Country** field to a blank area on your report canvas to create a map from **Data** pane.

Lab 03 - Data Visualization



Drag **Profit** field to the map. Power BI automatically adds bubbles representing the relative profit of each location to the map visual.



2.2.6 Dashboard

Power Bi allows users to create interactive dashboards to share information between visualizations.

For instance, we will use the **Slicer** to change the charts by date.

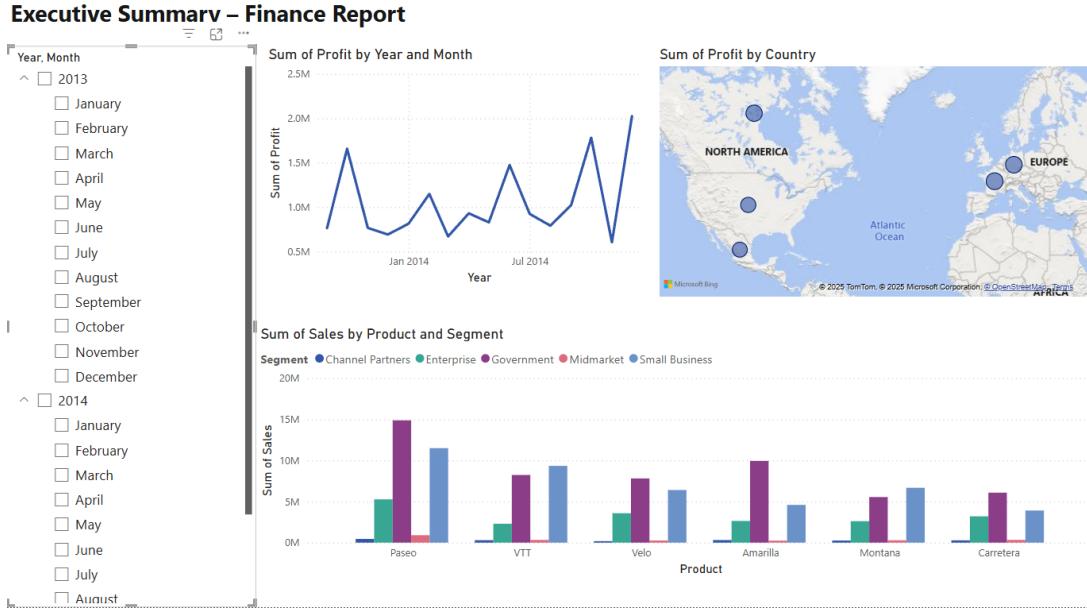


Figure 1: Before applying the filter

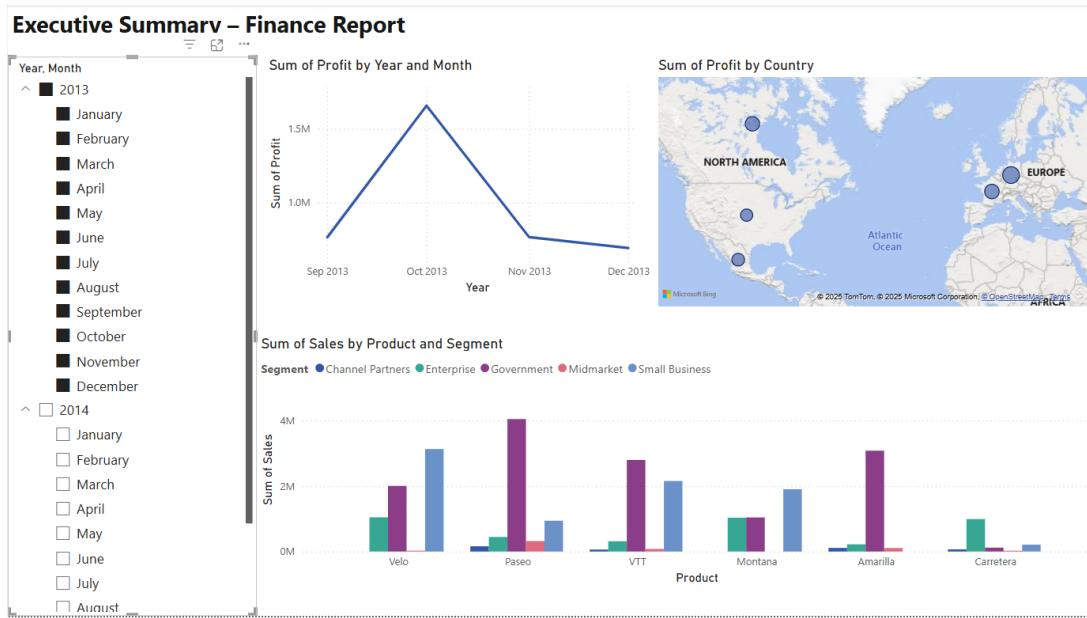
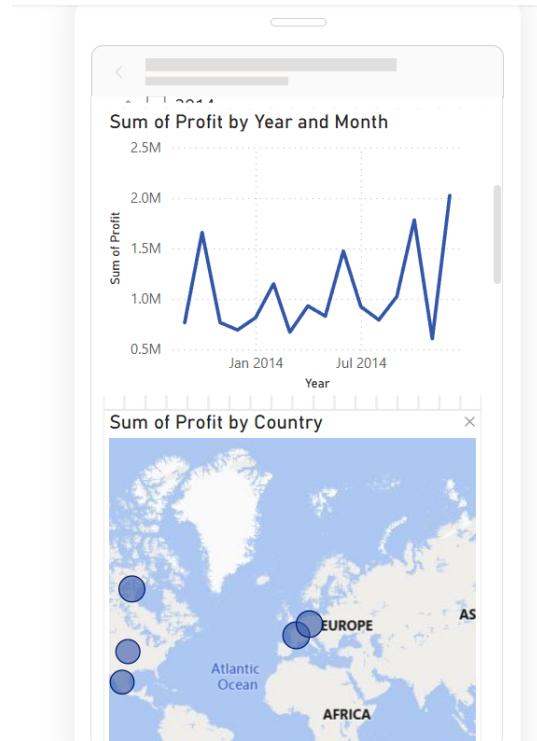


Figure 2: After applying the filter

2.2.7 Responsive devices

Power BI allows users view the dashboard on both Power BI desktop and Power BI mobile,

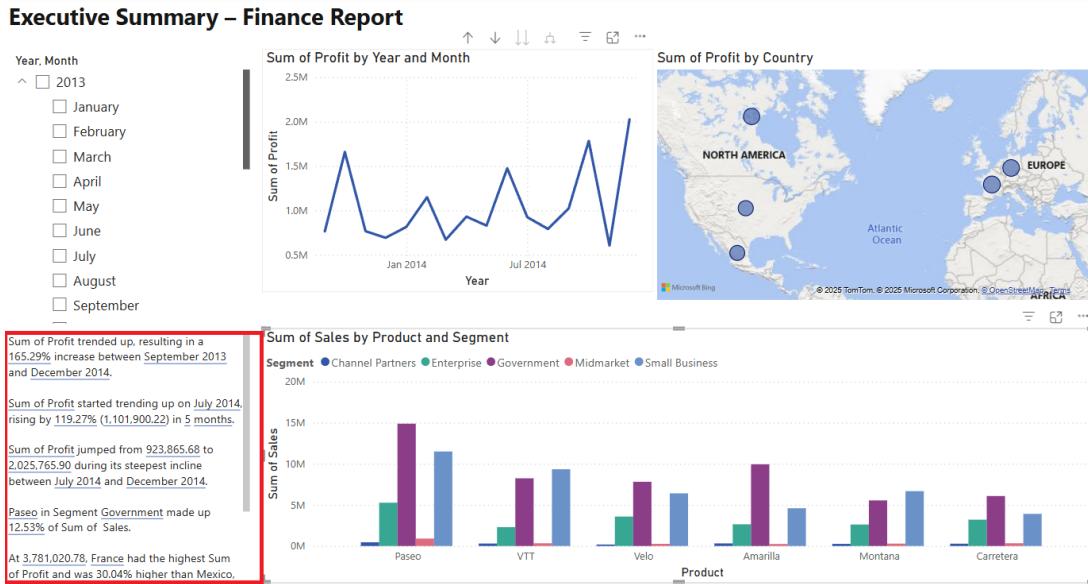


2.2.8 Auto-generated narrative

Narrative is a summary text for stakeholders who want to get insight from the dashboard. In Power BI, narrative can be generated automatically.

1. Select **Narrative** on **Insert** tab. Then click on **Custom** button.

2. Power BI automatically generates texts on the narrative field.



3 Main visualization task

3.1 Data collection

The data is collected on Kaggle about FIFA players dataset including stats of all players available in the game FIFA 22. The data has **CC0: Public Domain** license so that we are allowed to get the data there.

3.2 Data preprocessing

First, we removed index and url columns. Then we removed the columns that are no meaningful for further analysis based on our domain knowledge and data exploration.

Then, we checked for missing values. There are some columns which have high or low missing rate. For the columns of player's performance stats (pace, shooting, passing, dribbling, defending, physic), through exploration, we saw that the missing values in these columns are equal. We understood the reason why it has missing values among those columns. It is that the Goal Keeper players (GK) do not have these attributes. So we decided to fill these columns by 0 for GK players to keep this player position for analysis.

The remaining missing columns, if the missing values are below 1000, we removed the columns. If not, the columns just have below 100 missing records and we removed the rows that have missing values in these columns.

After that, we removed the columns which represent player's position as numeric type to reduce the width of the data for simplicity, since we already have club position and player positions columns.

Through checking, there are no duplicates in the dataset after having done all steps above.

Finally, we saved the preprocessed data into csv.

3.3 Data exploration

The cleaned dataset has 19165 rows and 58 columns. Each row represents a FIFA 22 player's record. Meanings of each column which are taken from the dataset source:

Column Name	Meaning
short_name	Player short name
player_positions	Player preferred positions
overall	Player current overall attribute
potential	Player potential overall attribute
value_eur	Player value (in EUR)
wage_eur	Player weekly wage (in EUR)
age	Player age
height_cm	Player height (in cm)
weight_kg	Player weight (in kg)
club_name	League name of the club
league_name	League rank of the club (e.g. English Premier League is 1, English League Championship is 2, etc.)
league_level	Player position in the club (e.g. SUB means substitute, RES means reserve)
club_position	Player position at the club
club_jersey_number	Player jersey number in the club
nationality_name	Player nationality
preferred_foot	Player preferred foot
weak_foot	Player weak foot attribute
skill_moves	Player skill moves attribute
international_reputation	Player international reputation attribute
pace	Player pace attribute
shooting	Player shooting attribute
passing	Player passing attribute
dribbling	Player dribbling attribute
defending	Player defending attribute
physic	Player physical attribute
attacking_crossing	Player crossing attribute
attacking_finishing	Player finishing attribute
attacking_heading_accuracy	Player heading accuracy attribute
attacking_short_passing	Player short passing attribute
attacking_volleys	Player volleys attribute
skill_dribbling	Player dribbling attribute
skill_curve	Player curve attribute
skill_fk_accuracy	Player free-kick accuracy attribute
skill_long_passing	Player long passing attribute
skill_ball_control	Player ball control attribute
movement_acceleration	Player acceleration attribute
movement_sprint_speed	Player sprint speed attribute
movement_agility	Player agility attribute
movement_reactions	Player reactions attribute
movement_balance	Player balance attribute
power_shot_power	Player shot power attribute
power_jumping	Player jumping attribute
power_stamina	Player stamina attribute
power_strength	Player strength attribute
power_long_shots	Player long shots attribute

Column Name	Meaning
mentality_aggression	Player aggression attribute
mentality_interceptions	Player interceptions attribute
mentality_positioning	Player positioning attribute
mentality_vision	Player vision attribute
mentality_penalties	Player penalties attribute
mentality_composure	Player composure attribute
defending_marking_awareness	Player marking awareness attribute
defending_standing_tackle	Player standing tackle attribute
goalkeeping_diving	Player GK diving attribute
goalkeeping_handling	Player GK handling attribute
goalkeeping_kicking	Player GK kicking attribute
goalkeeping_positioning	Player GK positioning attribute
goalkeeping_reflexes	Player GK reflexes attribute

3.4 Data visualization

3.4.1 Implementation Overview

For this dataset, we created 4 dashboards, including overview dashboard and 3 remaining dashboards for analysis.

Our team has 2 members so we derived 6 questions:

1. What is the player's important main stat (Passing, Shooting, Pace, Physic, Defending, Dribbling) that contributes the most to the Overall stat?
2. What are the differences in stats between the players that have the basic position (CF, CM, CB, GK) in football?
3. How might age-related player values impact transfer strategies for Manchester City and FC Barcelona, given their average wages and league?
4. What financial and competitive implications face English Premier League clubs versus German Bundesliga clubs, considering top player values and age trends?
5. How does the player's transfer value reflect their overall impact, and what influence could they have on the club's playing style?
6. How tactically adaptable is the player, and how effective are they in high-pressure or critical match situations?

About the dashboards, we use a wide range of dashboard types to ensure variety. Our team carefully examines various aspects, such as overall football performance, individual player statistics, player costs, and player comparisons. To achieve these objectives, we utilize a variety of graphs and charts for clear and effective data visualization including radar chart, bar chart, treemap, world map, funnel chart, donut chart and histogram.

For color usage, we use a football-inspired color palette, with green and blue as the dominant colors. For the colors, we prioritize using green for the background to provide a friendly look to the dashboard, aligning with the colors of football to make it easier for users to follow. The charts use shades of blue to stand out against the background for better visualization. This choice of football-themed colors is also user-friendly for individuals with red-green color blindness.

3.4.2 Dashboards and Analysis Objectives

I. Dashboard 0: Overview

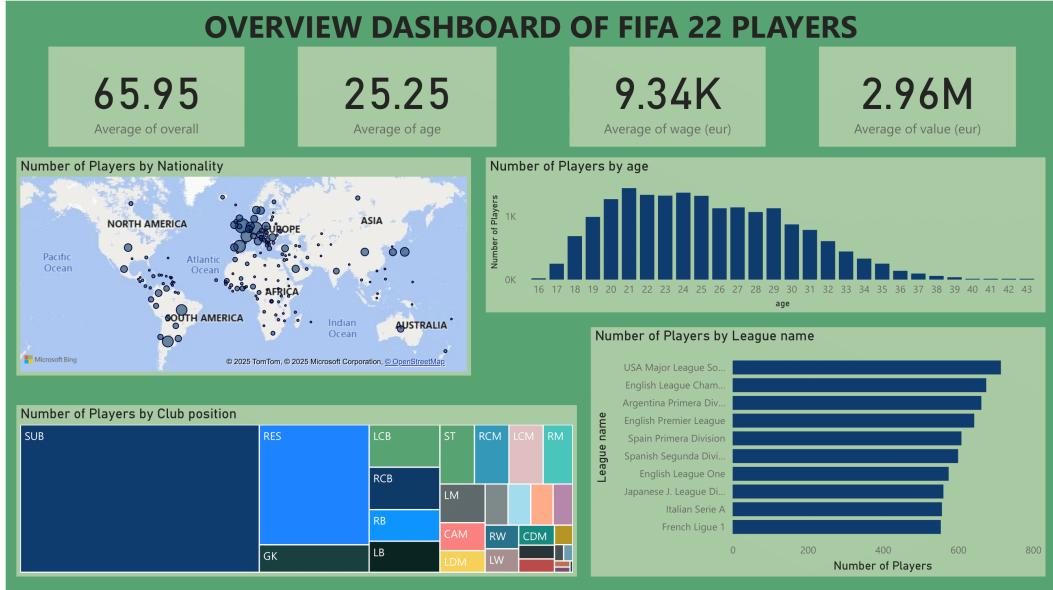


Figure 3: Overview Dashboard

For the overview dashboard, we mainly summarized the main points of the data knowledge. The data is mainly about FIFA 22 players. We created 4 cards to show players average overall stat, age, wage, and value, which are some basic attributes. Next, we concentrated on plotting the distribution of players by some attributes such as nationality, age, club position and league. For distribution of nationality, we used world filled map to see where the players mainly come from. For club position, we visualized by Treemap to see the difference between each positions. For age one, we plotted by column chart to easily witness the distribution. Finally, number of players by league name, we just plotted the top 10 leagues that contribute the most players in the dataset to highlight which league is the dominant.

Features used: short_name, age, nationality_name, league_name, club_position

II. Dashboard 1: Players Performance

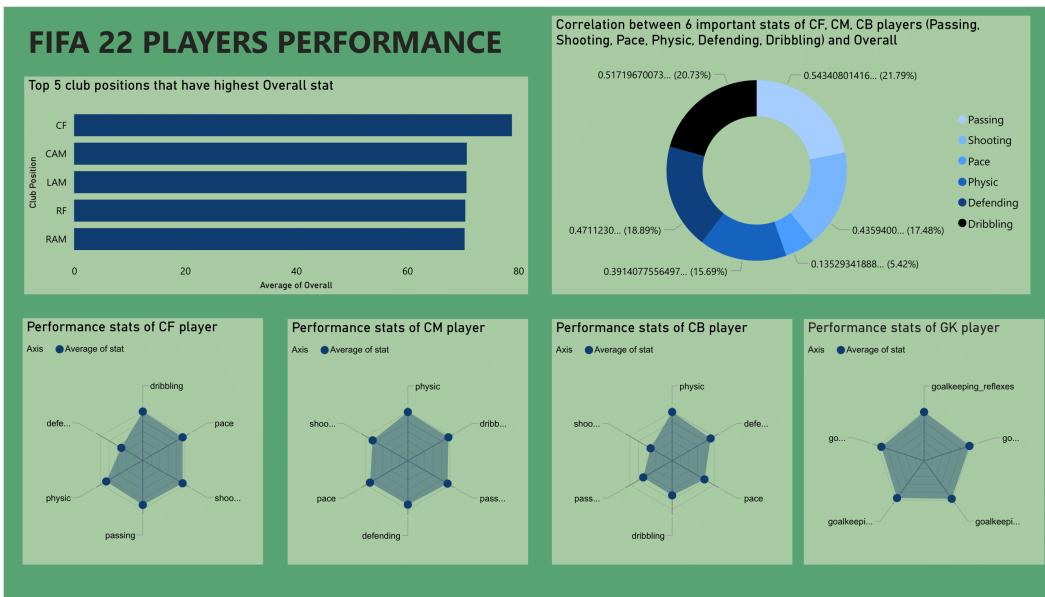


Figure 4: Players Performance Dashboard

For the first dashboard, we focused on analyzing players performance by their club positions, overall stat and important stats (Passing, Shooting, Pace, Physic, Defending, Dribbling). To visualize the highest performance by club positions, we used bar chart to identify the top 5 positions that have highest overall stat and witness their difference. We also had a donut chart to see which attribute contributes the most to main overall stat. The metric is calculated by Pearson correlation. We calculated the correlation between correlation and each player's important stat. Then, we dive into analyzing those important stats among the most vital positions based on football domain knowledge, that are Center Forward (CF), Central Midfielder (CM), Central Back (CB) and Goal Keeper (GK). To do that we used 4 radar chart which is suitable for visualizing many attributes with the same scale and easily to compare to other radar charts. But instead of using important stats like Passing, Shooting, Pace, Physic, Defending, Dribbling for all positions, we just used them for CF, CM and CB players. For GK players, we chose reflexes, diving, positioning, handling and kicking.

Features used: short_name, club_position, passing, dribbling, pace, shooting, physic, defending, goalkeeper_diving, goalkeeper_reflexes, goalkeeper_kicking, goalkeeper_handling, goalkeeper_positioning

Question 1: What is the player's important main stat (Passing, Shooting, Pace, Physic, Defending, Dribbling) that contributes the most to the Overall stat?

Benefit: Identifying the main stat that most influences Overall helps clubs and analysts prioritize player attributes for recruitment and development.

To answer this question, we look at the donut chart named "Correlation between 6 important stats of CF, CM, CB players (Passing, Shooting, Pace, Physic, Defending, Dribbling) and Overall". The values are calculated by Pearson correlation. From the chart, we can see that Passing contributes the most to Overall stat, while Pace contributes the least. Passing strongly impacts a player's rating—especially for CM who rely on distribution and vision. While Pace is valuable in real matches for quick breaks or covering ground, it's not heavily weighted in the Overall formula for these central positions.

Question 2: What are the differences in stats between the players that have the basic position (CF, CM, CB, GK) in football?

Benefit: Understanding stat differences by position allows for better role-specific training, recruit-

ment, and tactical planning.

To answer this question, we witness the 4 radar charts.

1. For CF players, they have high dribbling, pace and shooting stats, which are necessary for their efficient scoring skill. While their defending skill is the lowest because they mainly play at the higher position on field.
2. CM players may have lower individual stats, but their attributes are well-balanced. CM needs decent Passing for distribution, Dribbling to maintain control, Defending and Physic for ball recovery, and even some Shooting ability when pushing forward as playing at the center of field.
3. CB have a stat distribution that is almost the opposite of CF, excelling in Defending and Physic, while generally having lower Shooting and Dribbling. They are essential for breaking up attacks, winning aerial duels, and maintaining defensive structure.
4. GK players usually have a well-balanced overall rating across key goalkeeper stats. These include stats like Diving, Handling, Reflexes, Kicking, and Positioning. All these skills are important because a goalkeeper needs to be good at different things like stopping shots, catching crosses, making quick saves, and distributing the ball.

Those are the difference in stats between the players with different main positions.

III. Dashboard 2: Players Cost

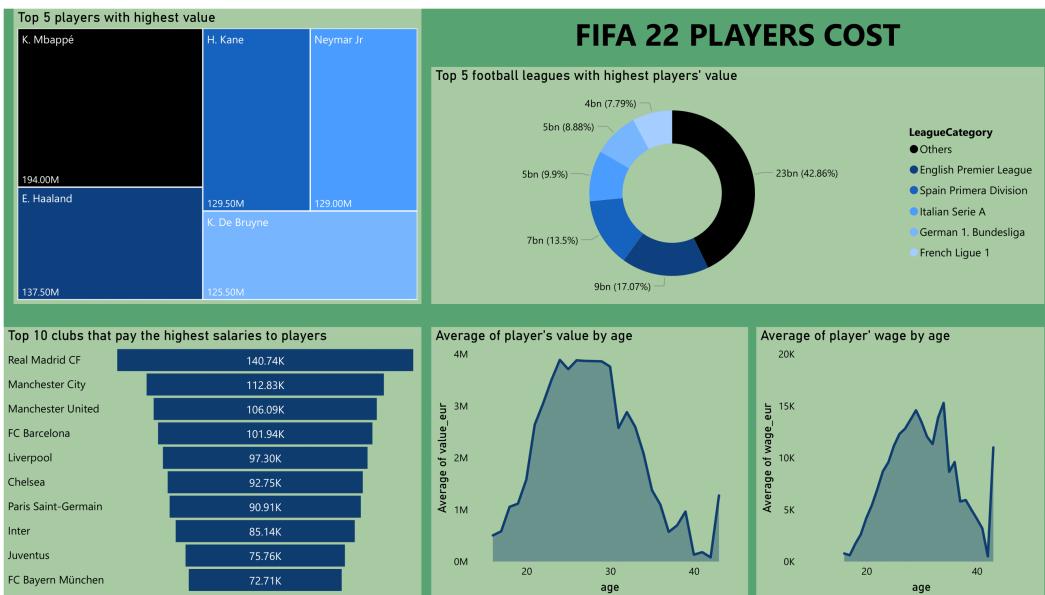


Figure 5: Players Cost Dashboard

In the second dashboard, we concentrated on considering player's wage, value and some contributing factors. We created Treemap to visualize top 5 players that have highest value and easily check their difference. We also used donut chart for simply identifying the league that contributes the highest summary of player's value. And to check top 10 clubs that have highest player's wage we used funnel chart to visually highlight the difference in wages across clubs. Finally, we used line chart to easily check the distribution of wage and value by age.

Features used: short_name, age, wage_eur, value_eur, league_name, club_name

Question 3: How might age-related player values impact transfer strategies for Manchester City and FC Barcelona, given their average wages and league?

Benefit: Analyzing age-related player values aids Manchester City and FC Barcelona in optimizing transfer strategies and wage management aligned with their league dynamics.

To answer this question, we look at the peak of each plot since the clubs we are analyzing belong to the top ones (Manchester City and FC Barcelona). Average of player's value by age line chart that shows player value peaks at 25-30 years (around 3.5M EUR). Wages peak at around 29 and at 34 (15K EUR). Witnessing the donut chart, we can see that where those clubs belongs to, English Premier League (Manchester City) contributes the highest value in the total summary of value, while Spain Primera Division (FC Barcelona), despite lower value than Premier League, but still contributes a large proportion. All of this show that Manchester City can buy peak-age players and sell older ones, for example Erling Haaland as shown in the treemap, while FC Barcelona do the same but be more careful with money.

Question 4: What financial and competitive implications face English Premier League clubs versus German 1. Bundesliga clubs, considering top player values and age trends?

Benefit: Comparing Premier League and Bundesliga clubs reveals financial and competitive challenges shaped by age trends and top player market values.

To answer this question, we check the summary of total player's value donut chart. It shows the English Premier League at 24bn EUR (43.72%) and German 1. Bundesliga at 7bn EUR (13.33%). As mentioned from the third question, player value peaks at 25-30 years, wage is also around that range. We look up to the clubs belong to those league, we see that most of the top value clubs are from Premier League clubs. We can answer that Premier League clubs, with 24bn EUR, buy peak-age players (25-30) like Haaland, boosting competitiveness but risking financial issues. German 1. Bundesliga clubs, with 7bn EUR, focus on younger players to save money but struggle as talent leaves for richer leagues.

IV. Dashboard 3: Players Comparison



Figure 6: Players Comparison Dashboard

The third dashboard was designed to compare two optional players by players' short name, club name and league name that the viewer's input. The dashboard has two boxes on the left and right

side, containing bar charts to visualize each player's attribute. It also contains two boxes between those two above for visualizing the viewer input, as well as display age, preferred foot, position, nationality, club name, league name, and overall. And two radar charts to compare the main important stats of two players.

In this context, for simplicity we just make only the case, which is comparing the players that have highest transferring value (Mbappe) with the highest quantity of fans player that play at the same position (Ronaldo). We will analyze these two players based on their attributes.

Features used: short_name, club_name, league_name, age, club_postion, preferred_foot, overall, attacking_crossing, attacking_finishing, attacking_heading_accuracy, attacking_short_passing, attacking_volleys, skill_dribbling, skill_curve, skill_curve, skill_fk_accuracy, skill_long_passing, skill_ball_control, movement_acceleration, movement_sprint_speed, movement_agility, movement_reactions, movement_balance, power_shot_power, power_jumping, power_stamina, power_strength, power_long_shots, mentality_aggression, mentality_interceptions, mentality_positioning, mentality_vision, mentality_penalties, mentality_composure, goalkeeping_positioning

Question 5: How does the player's transfer value reflect their overall impact, and what influence could they have on the club's playing style?

Benefit: Evaluating a player's transfer value in relation to their overall impact helps clubs assess their influence on team performance and playing style.

To answer this question, we check the age difference (Mbappé at 24 and Ronaldo at 36) and the disparities in sprint speed, stamina, and dribbling, what factors explain Mbappe's higher transfer value (160M eur) compared to Ronaldo's (20M eur). Mbappe is younger (24), so he can play at a high level for longer. His sprint speed (97) and stamina (88) are better than Ronaldo's (85 and 77), meaning he is faster and lasts longer in games. His dribbling (92) is also higher than Ronaldo's (88), showing he's great at creating chances. This makes him worth 160 million euros, while Ronaldo, nearing the end of his career, is only valued at 20 million euros. Mbappe's speed and dribbling help the team play fast, attacking quickly and breaking through defenses. His team can switch styles easily because he's good at passing too (vision 82). Ronaldo's finishing (95) and positioning (95) mean they focus on scoring from crosses or set pieces (heading accuracy 90), but they are less flexible since he is slower and tires more easily.

Question 6: How tactically adaptable is the player, and how effective are they in high-pressure or critical match situations?

Benefit: Measuring a player's tactical adaptability and clutch performance ensures reliability in high-stakes matches and informs tactical decisions.

To answer this question, we consider the differences in vision, long shots, and heading accuracy in the dashboard, how do Mbappe and Ronaldo adapt to varying tactical systems, and which of the two holds a greater advantage in delivering decisive moments during high-stakes matches in Ligue 1 and the Premier League. Mbappe's vision (82) helps him to set up plays and work well with teammates, fitting PSG's fast and creative style. Ronaldo's vision (76) is lower, so he is more about finishing chances (finishing 95), suiting Manchester United's plan to get the ball to him for goals. Ronaldo's heading accuracy (90) makes him strong in the air, perfect for scoring in tight Premier League games. His long shots (93) also help him score from far out. Mbappe's "long shots" (82) are decent, and his vision (82) lets him create chances, making him clutch in Ligue 1's quick games. Mbappe might edge out slightly because he can do more things, while Ronaldo needs good passes to shine.

References

- [1] What is Power BI?
- [2] Power BI functions and examples