Data Visualization Presentation

Group 1
Peter Vanderhook
Tamim Hemat
Daryush Balsara

Where did our dataset come from?

Tableau provides free public sample data

Have sources for all resources provided

Our specific dataset was sourced from data.world

It contains data collected from 2018 FIFA WC rosters

See our references for more info

2018 FIFA World Cup Rosters

Goals, caps, club, and date of birth for players on 2018 FIFA World Cup rosters. Source: data.world

Dataset (xlsx)

What does our dataset look like?

Our dataset was provided in XLS format

Used google drive to convert to CSV

CSV file's first row contains column names

Type contained two unique values, each of which had duplicate rows

Players were separated into 8 groups

Caps means times the player was present for a game

Type	Team	Group	Position	Name	DOB	Caps	Goals	Country and Club
Age	Egypt	A	DF	Ahmed Elmohamady	1987-09-09	76	2	England Aston Villa
Age	Egypt	A	DF	Ahmed Fathy	1984-11-10	124	3	Egypt Al Ahly
Age	Egypt	A	DF	Ahmed Hegazi	1991-01-25	43	1	England West Bromwi...
Age	Egypt	A	DF	Ali Gabr	1993-07-28	20	1	England West Bromwi...
Age	Egypt	A	DF	Ayman Ashraf	1991-04-09	4	0	Egypt Al Ahly
Age	Egypt	A	DF	Mahmoud Hamdy	1995-06-01	0	0	Egypt Zamalek
Age	Egypt	A	DF	Mohamed Abdel-Shafy	1985-07-01	49	1	Saudi Arabia Al Fateh

How did we clean our data?

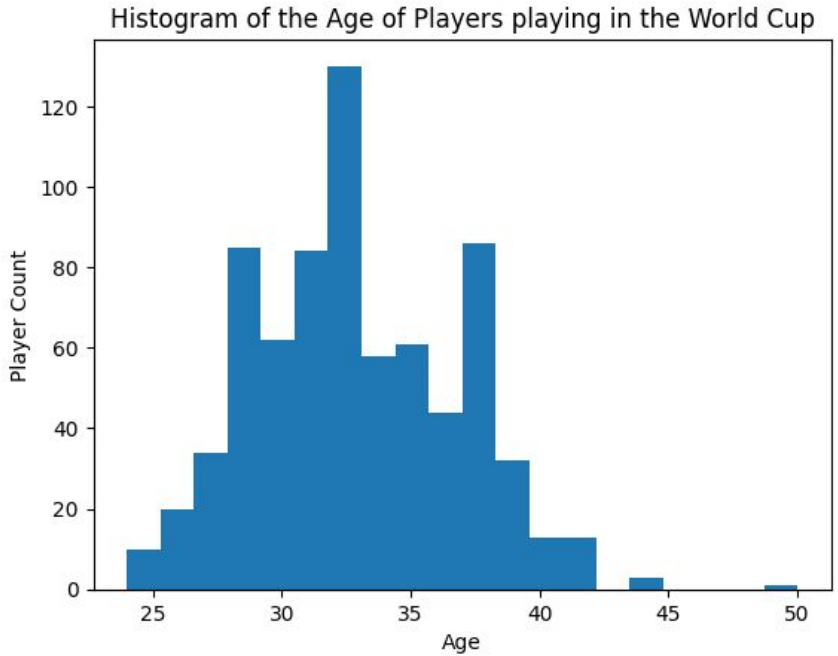
```
import pandas as pd

#   Load the original CSV File
df = pd.read_csv('worldcup.csv')

#   Remove the rows where types = 'Caps'
df = df[df['Type'] != 'Caps']

#   Now that duplicate rows are removed, remove the entire type column
df = df.drop(columns=["Type"])

#   Name the first column of index numbers.
df.index.name = 'id'

#   Save the dataframe to a csv file
df.to_csv('worldcup_clean.csv')
```

Some Visualization with Jupyter

We are provided only with DOB

Can calculate age by *current_year - DOB_year*

Youngest player was 24, oldest was 50



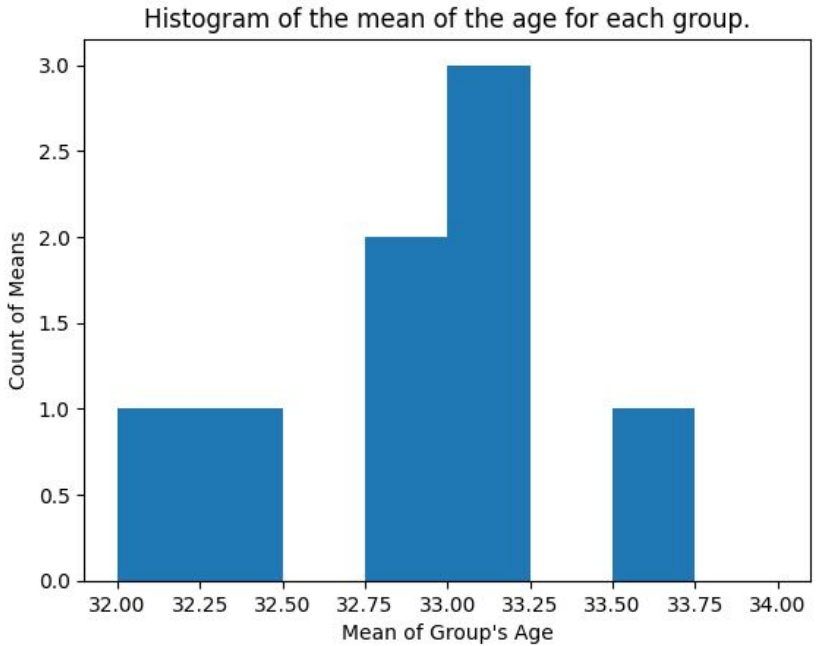
More Visualization on Jupyter

Here we plot a histogram for values from each group

We calculated age and stored in a dictionary

The dictionary uses the Group as a key

We can see the average age for each group is similar



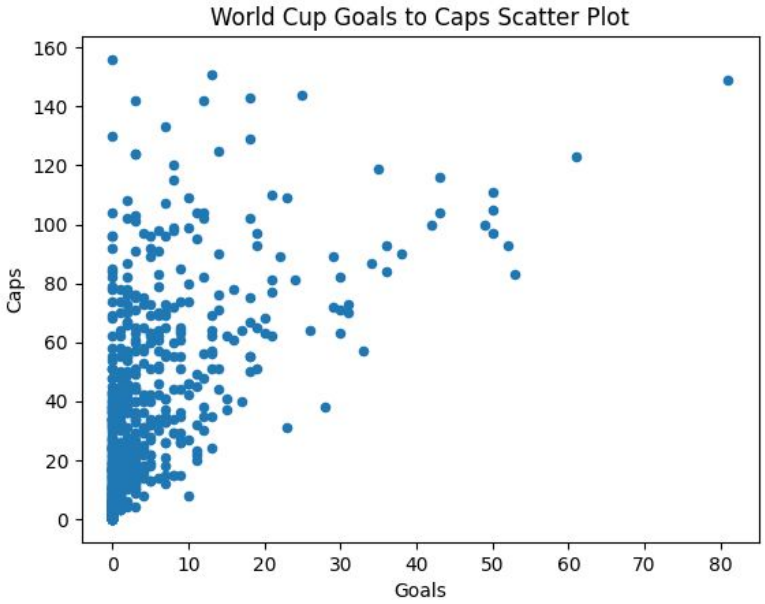
Final Visualization on Jupyter

We are provided with Caps and Goals

Scatter plot of player goals(x) and caps(y)

Lots of players had few caps or goals

One player had 160 caps and 0 goals (rip)



Insights from these Stats

Lots of players had few caps and goals

Despite wide range of ages, most are early 30s

Each group participating has roughly the same mean age for its players.

Tableau Portion

Now it’s time to get interactive

Questions?

References

Data
https://public.tableau.com/app/resources/sample-data
https://data.world/sportsvizsunday/sports-viz-sundays-2018/workspace/file?filename=World+Cup+2018+Squads.xlsx

Tools
https://drive.google.com/
https://jupyter.org/

Docs
https://numpy.org/doc/
https://pandas.pydata.org/docs/