**FIFA WORLD CUP**

Course: Work Integrated Learning

Group Members: Harshil Shah

Date: 14/12/2022

**TABLE OF CONTENTS**

Page

1.0 Introduction 3

1.1 Problem statement 3

1.2 Data 3

1.3 Tasks and responsibilities 3

2.0 Methodology 3

2.1 Data Cleaning 4

2.1.1 Data Transformation 4

2.1.1.1 Analysis using PowerBi 5

3.0 Results 7

4.0 CONCLUSIONS 8

5.0 REFERENCES 8

# Introduction

As we all know about “FIFA WORLD CUP”, the world waits for it and I have used that data for my project to analyse the data.

## Problem statement

To know about how popular and demanding this game gets day by day.

## Data

Analysing the information and breaking it down into smaller parts to get best results out of it. This dataset is used from Kaggle.

## Tasks and responsibilities

reference Table 1 here

**Table 1: example: list of tasks**

|  |  |
| --- | --- |
| Student name | Tasks |
| Harshil Shah | Everything |

# Methodology

Reading the data:

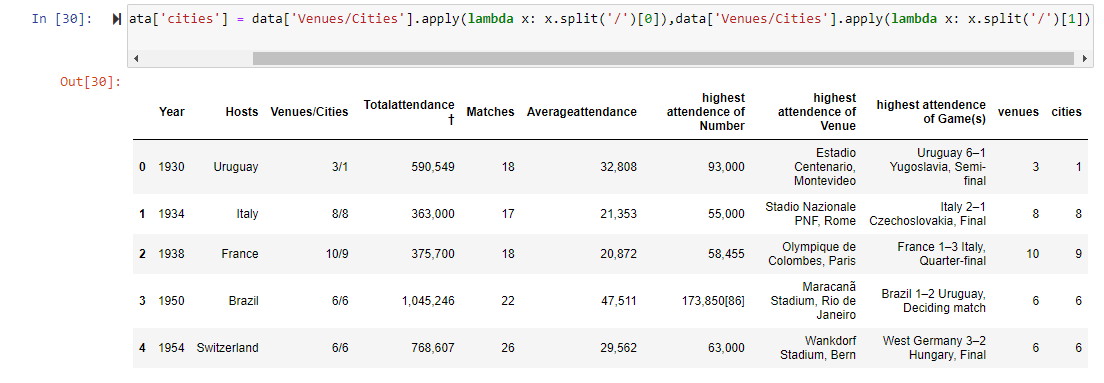
Importing required libraries for reading data and getting the structure of data. We can see all the data which almost needs a complete cleaning. Furthermore, I will be explaining about the next steps taken.

Graphical user interface

Description automatically generated with medium confidence

## Data Cleaning

This data was completely cleaned and transformed so that we can get a clear picture about how and what we are working on. Firstly, venues/cities were separated and made separate columns to get better results, then “,” were removed between the numbers to get accurate information.



### Data Transformation

Total attendance column is being cleaned by removing the “,” using string function. Then we changed average attendance by removing “,” between the numbers. After this we have highest attendance of number in which we need to change the “,” as well. Lastly, we removed all the “,” so that our data is cleaned and ready to use.

Graphical user interface

Description automatically generated with medium confidence

#### Analysis using PowerBi

Chart, line chart

Description automatically generated

This shows us the difference between how people are enjoying the craze of FIFA world cup as you can see average attendance has been increasing every year starting from 1920 its been almost 100 years and touching new heights every year.

Chart, bar chart

Description automatically generated

The above figure shows us how many matches played by the host country every year. The highest is 60 matches and its not only by 1 host but it includes 4 hosts and the names are Germany, Russia, South Africa and South Korea Japan respectively.

Chart, pie chart

Description automatically generated

This pie charts tells about the winners of every year from 1930 to 2020 and as you can see Brazil comes at top with 14.27% which is a very good score and shows how strong they are in this game. The second is West Germany which is very close with 14.26%. Lastly, comes England with 4.74%.

# Results

Chart, pie chart

Description automatically generated

This shows the attendance of Game type which clearly shows us a winner with almost 47.6% is the Final match of world cup which is the most enjoyed and loved match type by audience. Following with Group stage with 23.8%.

# CONCLUSIONS

Chart, pie chart

Description automatically generated

In conclusion I would like to show you one of the most important analysis and it will be the last analysis of this report which is the team people enjoy the most during the FIFA World Cup and as you can see that there is a tie between 2 teams which are WestGermany and Brazil respectively. They have 14.29% each as a audience control in the whole stadium. This is one of the great and best thing for the world.

# REFERENCES

1. <https://www.kaggle.com/datasets/swaptr/fifa-world-cup-2022-match-data>
2. <https://plotly.com/python/plotly-express/>
3. https://seaborn.pydata.org/generated/seaborn.barplot.html.