FIFA WORLD CUP DATA ANALYSIS

Shahaan Kaushik
Thapar Institute of Engineering and Technology

TABLE OF CONTENT

INTRODUCTION

Data Analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision- making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains.

The analytics team of the Football World Cup association and FIFA association anywhere in the world would love to check our through data analysis of each and every match leading to a well-organized and fruitful information. My analysis contains data on host teams, all stadiums, most supported team and teams which are most successful.

DASHBOARD


FIFA World Cup
Analysis Dashboard
1.click here to viewdifference in golas dataset and data cleaning flow chart
3.click here to view total attendance dataset
5.click here to view no. of people came country wise dataset
7.click here to view Maximum worldcup winners
2.click here to view Total attendance dataset
4.click here to view overall attendance city wise dataset
6.click here to view Goals scored by teams data
8.click here to view top 3 teams dataset
Total Attendance
Total Winners
Total crowd of people
Top 3 Teams
Overall Goals
Stage
Final
First round
Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group A
Group B
Group C
Group D
Group E
Group F
Group G
Group H
Match for th...
Play-off for ...
Preliminary ...
Quarter-finals
Year
1930
1934
1938
1950
1954
1958
1962
1966
1970
1974
1978
1982
1986
1990
1994
1998
2002
2006
2010
2014
match wise attendance

SCOPE OF ANALYSIS

The system wants to see and analyze the sales trend month-wise and product-wise and work upon the lagging segments and outperforming employees accordingly. The Analytics team also wants to create analyze the database in depth to help the super store grow exponentially. The Analytics team wishes to answer the following objectives: -

- Half time goals scored by the home and away team year wise.
- No of goals scored in finals until now.
- Analyzing the total attendance of crowd in Group matches, semifinals and Finals.
- Overall Attendance for a particular city until now.
- No. of people came to view match for particular team.
- Country winning the maximum World Cups.
- Goals scored by home teams and away team
- Analyzing difference between total goals scored and total goals conceded by teams

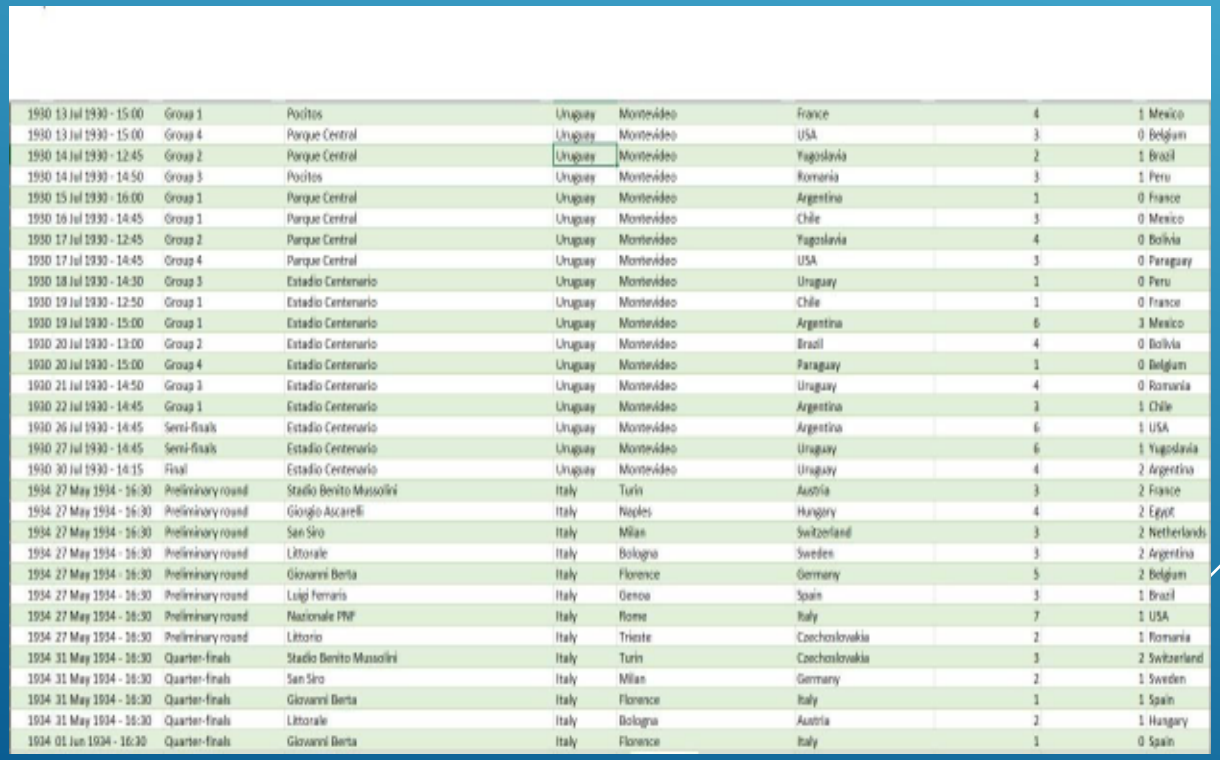
Aim of this project is to answer the above objectives in the form of visualization by creating a dashboard to convey the answers effectively and efficiently.

STEP1.REMOVING BLANK CELLS FROM THE DATASET
STEP2.REMOVING COLUMNS WHICH ARE NOT PROPERLY DEFINED OR NOT CRUCIAL TO OUR ANALYSIS

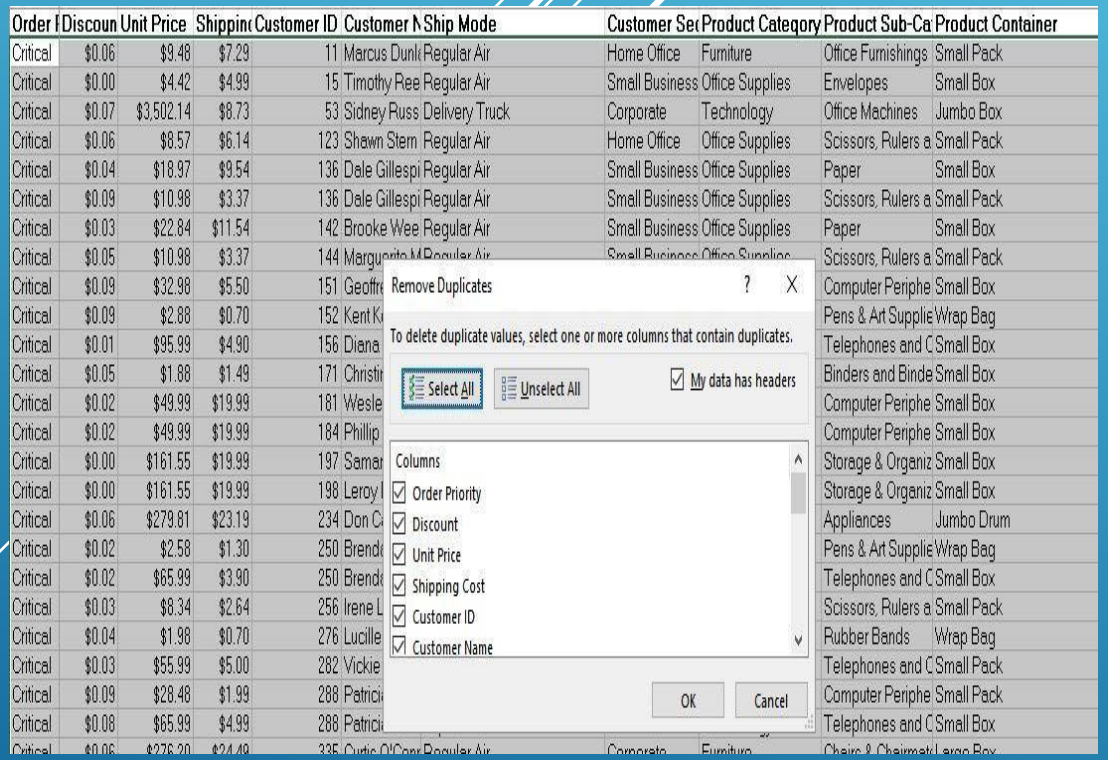
Step3. Giving proper and appropriate column names.

Step4. Removing Duplicate Values

Step5. Improvising Proper Data Formatting



1930 13 Jul 1930 - 15:00	Group 1	Pocitos	Uruguay	Montevideo	France	4	1 Mexico
1930 13 Jul 1930 - 15:00	Group 4	Parque Central	Uruguay	Montevideo	USA	3	0 Belgium
1930 14 Jul 1930 - 12:45	Group 2	Parque Central	Uruguay	Montevideo	Yugoslavia	2	1 Brazil
1930 14 Jul 1930 - 14:50	Group 3	Pocitos	Uruguay	Montevideo	Romania	3	1 Peru
1930 15 Jul 1930 - 16:00	Group 1	Parque Central	Uruguay	Montevideo	Argentina	1	0 France
1930 16 Jul 1930 - 14:45	Group 1	Parque Central	Uruguay	Montevideo	Chile	3	0 Mexico
1930 17 Jul 1930 - 12:45	Group 2	Parque Central	Uruguay	Montevideo	Yugoslavia	4	0 Bolivia
1930 17 Jul 1930 - 14:45	Group 4	Parque Central	Uruguay	Montevideo	USA	3	0 Paraguay
1930 18 Jul 1930 - 14:30	Group 3	Estadio Centenario	Uruguay	Montevideo	Uruguay	1	0 Peru
1930 19 Jul 1930 - 12:50	Group 1	Estadio Centenario	Uruguay	Montevideo	Chile	1	0 France
1930 19 Jul 1930 - 15:00	Group 1	Estadio Centenario	Uruguay	Montevideo	Argentina	6	3 Mexico
1930 20 Jul 1930 - 13:00	Group 2	Estadio Centenario	Uruguay	Montevideo	Brazil	4	0 Bolivia
1930 20 Jul 1930 - 15:00	Group 4	Estadio Centenario	Uruguay	Montevideo	Paraguay	1	0 Belgium
1930 21 Jul 1930 - 14:50	Group 3	Estadio Centenario	Uruguay	Montevideo	Uruguay	4	0 Romania
1930 22 Jul 1930 - 14:45	Group 1	Estadio Centenario	Uruguay	Montevideo	Argentina	3	1 Chile
1930 26 Jul 1930 - 14:45	Semi-finals	Estadio Centenario	Uruguay	Montevideo	Argentina	6	1 USA
1930 27 Jul 1930 - 14:45	Semi-finals	Estadio Centenario	Uruguay	Montevideo	Uruguay	6	1 Yugoslavia
1930 30 Jul 1930 - 14:15	Final	Estadio Centenario	Uruguay	Montevideo	Uruguay	4	2 Argentina
1934 27 May 1934 - 16:30	Preliminary round	Stadio Benito Mussolini	Italy	Turin	Austria	3	2 France
1934 27 May 1934 - 16:30	Preliminary round	Giorgio Ascarelli	Italy	Naples	Hungary	4	2 Egypt
1934 27 May 1934 - 16:30	Preliminary round	San Siro	Italy	Milan	Switzerland	3	2 Netherlands
1934 27 May 1934 - 16:30	Preliminary round	Littoriale	Italy	Bologna	Sweden	3	2 Argentina
1934 27 May 1934 - 16:30	Preliminary round	Giovanni Berta	Italy	Florence	Germany	5	2 Belgium
1934 27 May 1934 - 16:30	Preliminary round	Luigi Ferraris	Italy	Genoa	Spain	3	1 Brazil
1934 27 May 1934 - 16:30	Preliminary round	Nazionale PNF	Italy	Rome	Italy	7	1 USA
1934 27 May 1934 - 16:30	Preliminary round	Littorio	Italy	Trieste	Czechoslovakia	2	1 Romania
1934 31 May 1934 - 16:30	Quarter-finals	Stadio Benito Mussolini	Italy	Turin	Czechoslovakia	3	2 Switzerland
1934 31 May 1934 - 16:30	Quarter-finals	San Siro	Italy	Milan	Germany	2	1 Sweden
1934 31 May 1934 - 16:30	Quarter-finals	Giovanni Berta	Italy	Florence	Italy	1	1 Spain
1934 31 May 1934 - 16:30	Quarter-finals	Littoriale	Italy	Bologna	Austria	2	1 Hungary
1934 01 Jun 1934 - 16:30	Quarter-finals	Giovanni Berta	Italy	Florence	Italy	1	0 Spain



Order I	Discoun	Unit Price	Shippin	Customer ID	Customer N	Ship Mode	Customer Se	Product Category	Product Sub-Ca	Product Container
Critical	$0.06	$9.48	$7.29	11	Marcus Dunl	Regular Air	Home Office	Furniture	Office Furnishings	Small Pack
Critical	$0.00	$4.42	$4.99	15	Timothy Ree	Regular Air	Small Business	Office Supplies	Envelopes	Small Box
Critical	$0.07	$3,502.14	$8.73	53	Sidney Russ	Delivery Truck	Corporate	Technology	Office Machines	Jumbo Box
Critical	$0.06	$8.57	$6.14	123	Shawn Stern	Regular Air	Home Office	Office Supplies	Scissors, Rulers a	Small Pack
Critical	$0.04	$18.97	$9.54	136	Dale Gillespi	Regular Air	Small Business	Office Supplies	Paper	Small Box
Critical	$0.09	$10.98	$3.37	136	Dale Gillespi	Regular Air	Small Business	Office Supplies	Scissors, Rulers a	Small Pack
Critical	$0.03	$22.84	$11.54	142	Brooke Wee	Regular Air	Small Business	Office Supplies	Paper	Small Box
Critical	$0.05	$10.98	$3.37	144	Marguerite M	Regular Air	Small Business	Office Supplies	Scissors, Rulers a	Small Pack
Critical	$0.09	$32.98	$5.50	151	Geoffre				Computer Periphe	Small Box
Critical	$0.09	$2.88	$0.70	152	Kent K				Pens & Art Supplie	Wrap Bag
Critical	$0.01	$95.99	$4.90	156	Diana				Telephones and C	Small Box
Critical	$0.05	$1.88	$1.49	171	Christi				Binders and Binde	Small Box
Critical	$0.02	$49.99	$19.99	181	Wesle				Computer Periphe	Small Box
Critical	$0.02	$49.99	$19.99	184	Phillip				Computer Periphe	Small Box
Critical	$0.00	$161.55	$19.99	197	Samar				Storage & Organiz	Small Box
Critical	$0.00	$161.55	$19.99	198	Leroy				Storage & Organiz	Small Box
Critical	$0.06	$279.81	$23.19	234	Don C				Appliances	Jumbo Drum
Critical	$0.02	$2.58	$1.30	250	Brend				Pens & Art Supplie	Wrap Bag
Critical	$0.02	$65.99	$3.90	250	Brend				Telephones and C	Small Box
Critical	$0.03	$8.34	$2.64	256	Irene L				Scissors, Rulers a	Small Pack
Critical	$0.04	$1.98	$0.70	276	Lucille				Rubber Bands	Wrap Bag
Critical	$0.03	$55.99	$5.00	282	Vickie				Telephones and C	Small Pack
Critical	$0.09	$28.48	$1.99	288	Patrici				Computer Periphe	Small Pack
Critical	$0.08	$65.99	$4.99	288	Patrici				Telephones and C	Small Box
Critical	$0.06	$276.20	$24.49	335	Curtis O'Conr	Regular Air	Corporate	Furniture	Chairs & Chairmats	Large Box

Remove Duplicates

To delete duplicate values, select one or more columns that contain duplicates.

Select All | Unselect All | My data has headers

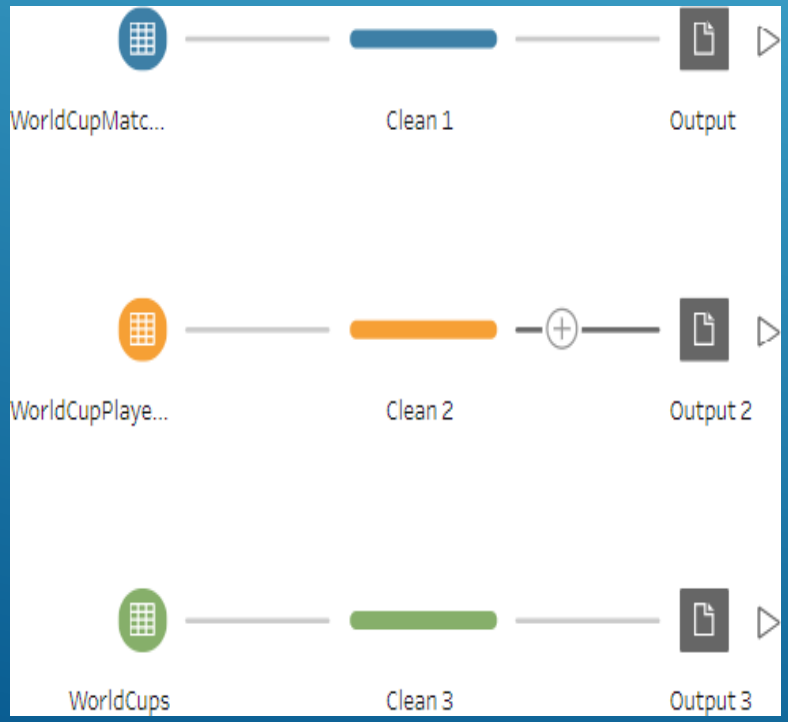
Columns

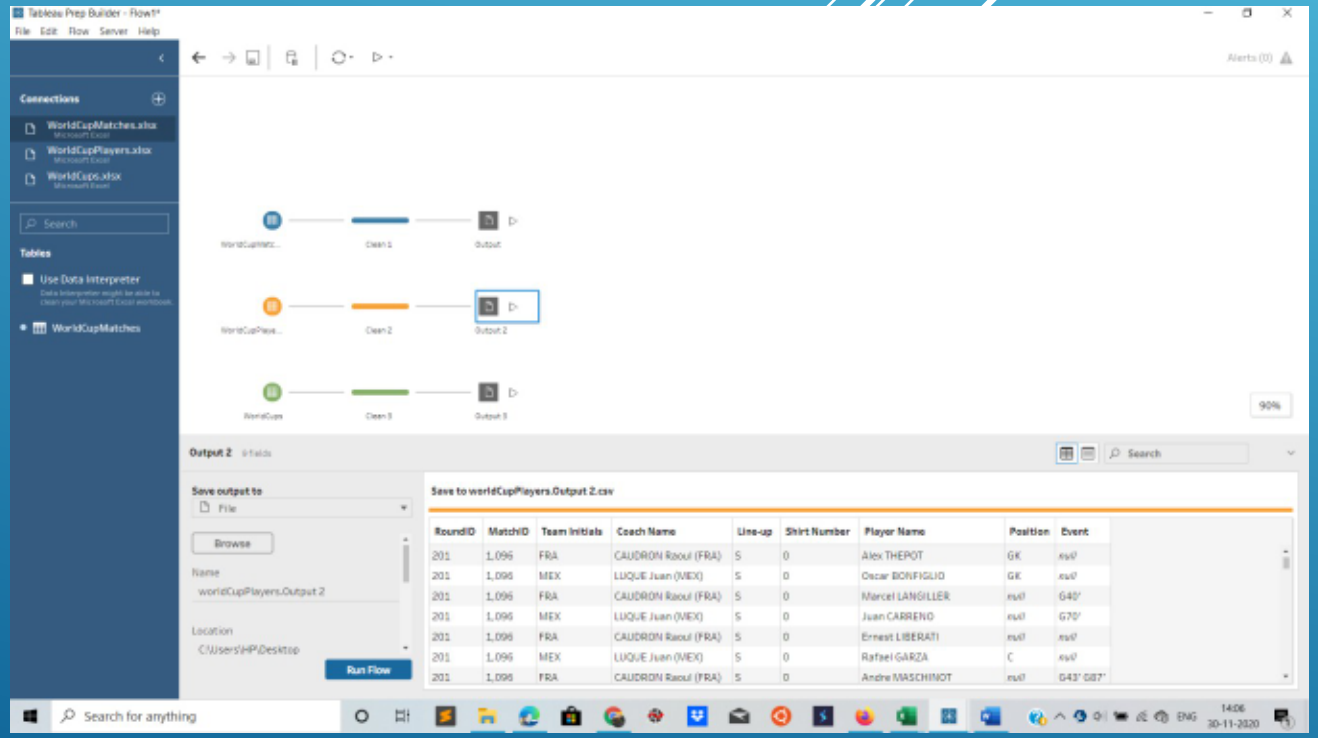
- Order Priority
- Discount
- Unit Price
- Shipping Cost
- Customer ID
- Customer Name

OK | Cancel

STEP6. EXCLUDING THE NULL VALUES FROM THE DATA.

We'll be using Tableau prep for this work as it'll make the work simple and faster because we might not know how many null values could be there in this huge data set. Tableau helps us doing one step cleaning with ease.





1.NO.OF GOALS SCORED IN FINALS UNTIL NOW

Description:

By calculating the current trend of the number of goals scored by home and away teams we can check who have strategical advantage and who is going under more pressure.

Specific function and requirements

As we can see total of 69 goals scored in finals out of which 1958 having the most of 7 goals and 1994 with least of 0 goals.

Stage	Final	
Year	Home Team Goals	Away Teams Goals
1930	4	2
1934	2	1
1938	4	2
1954	3	2
1958	5	2
1962	3	1
1966	4	2
1970	4	1
1974	1	2
1978	3	1
1982	3	1
1986	3	2
1990	1	0
1994	0	0
1998	0	3
2002	0	2
2006	1	1
2010	0	1
2014	2	0
Grand Total	43	26

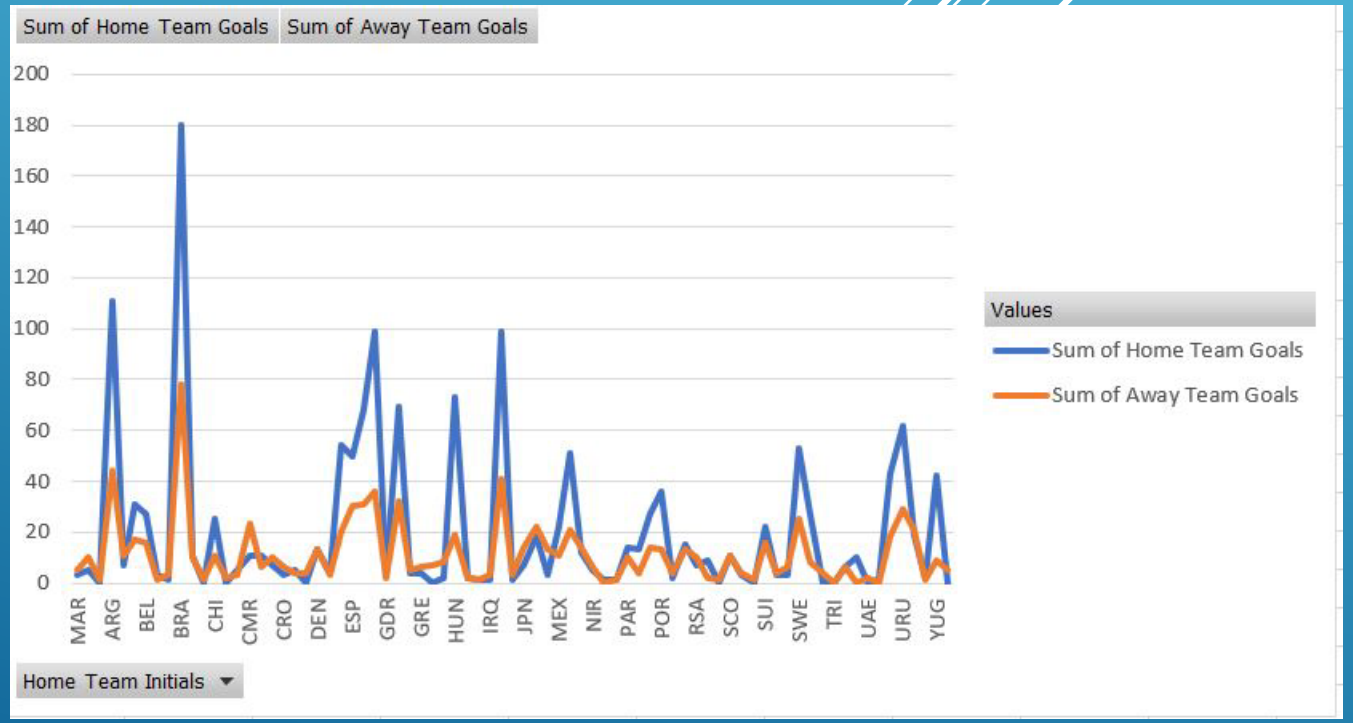
2.GOALS SCORED BY HOME TEAMS AND AWAY TEAMS

Describes the total goals scored by home and away teams and also top 3 teams scoing most goal difference.

Specific function and requirements:

We have to create a pivot table. No specific functions are used. Only a few snips of the data shown

Row La	Sum of Hc	Sum of Away Team Goals
MAR	3	5
ALG	5	10
ANG	0	1
ARG	111	44
AUS	7	11
AUT	31	17
BEL	27	16
BIH	3	1
BOL	1	3
BRA	180	78
BUL	11	10
CAN	0	1
CHI	25	11
CHN	0	2
CIV	5	3
CMR	11	23
COL	11	6
CRC	7	10
CRO	3	6
CUB	5	4
CZE	0	4
DEN	13	13
ECU	4	3
ENG	54	20
ESP	50	30
FRA	68	31
FRG	99	36
GDR	3	2
GER	69	32
GHA	4	5



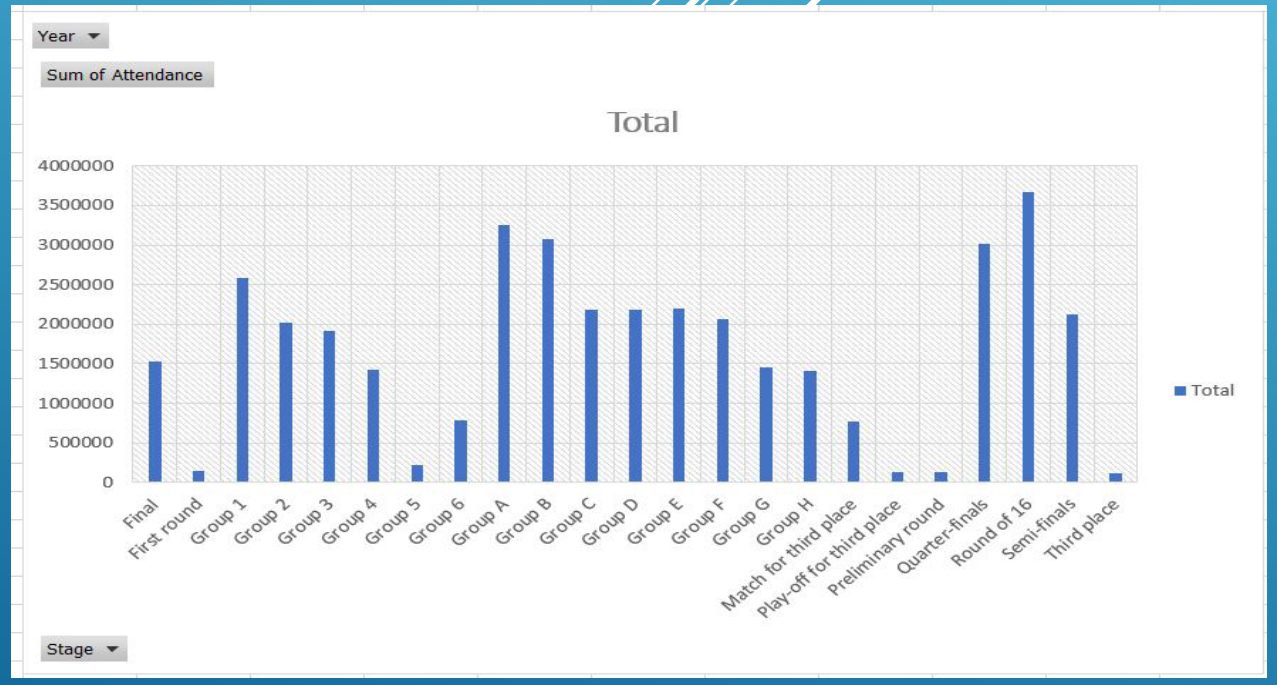
3.ANALYZING THE TOTAL ATTENDANCE OF CROWD IN GROUP MATCHES, SEMI-FINALS AND FINALS

Describes the total attendance based on the group A,B,C,D and semifinals and finals and determining which stadium to be chosen, approx. amount of people expected security required etc.

Specific function and requirements:

We have to create a pivot table. No specific functions are used. We then put the type *of match and sum of sales in the columns*.

Year	(All)
Sections	**Sum of Attendance**
Final	1527673
First round	145083
Group 1	2583172
Group 2	2020264
Group 3	1919199
Group 4	1425366
Group 5	212124
Group 6	787903
Group A	3259281
Group B	3082072
Group C	2184717
Group D	2180512
Group E	2192444
Group F	2056362
Group G	1455995
Group H	1402411
Match for third place	762718
Play-off for third place	136068
Preliminary round	135000
Quarter-finals	3016034
Round of 16	3664279
Semi-finals	2125920
Third place	115483
Grand Total	**38390080**



4.COUNTRY WINNING THE MAXIMUM WORLD CUP

Description:

Determines the country with maximum FIFA world cup trophy.

Specific function and requirements:

We have to create a pivot table. No specific functions are used.

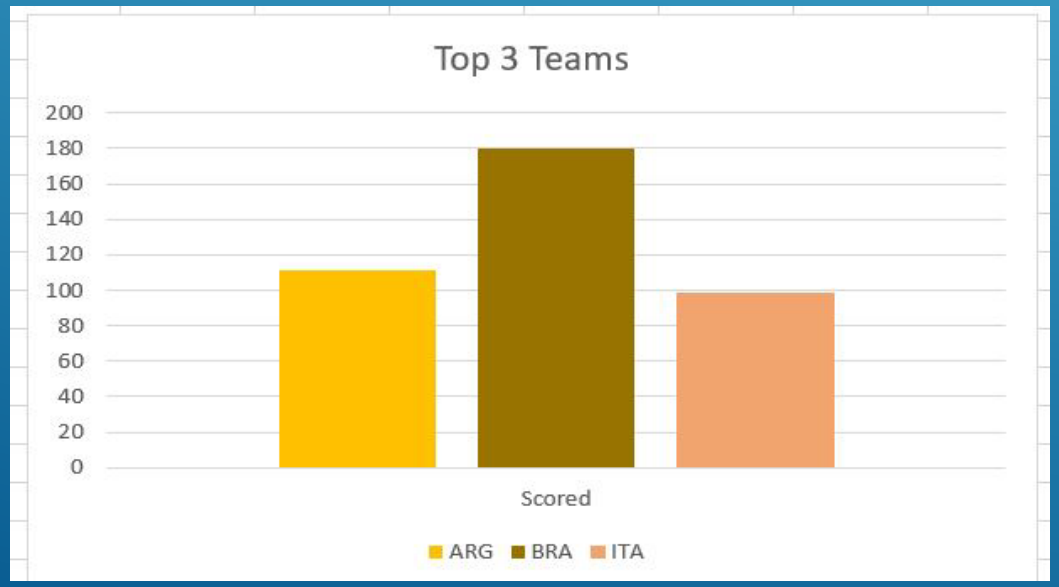
Country	Count of Winner
Argentina	1
Brazil	2
Chile	1
England	1
France	2
Germany	2
Italy	2
Korea/Japan	1
Mexico	2
South Africa	1
Spain	1
Sweden	1
Switzerland	1
Uruguay	1
USA	1
Grand Total	**20**

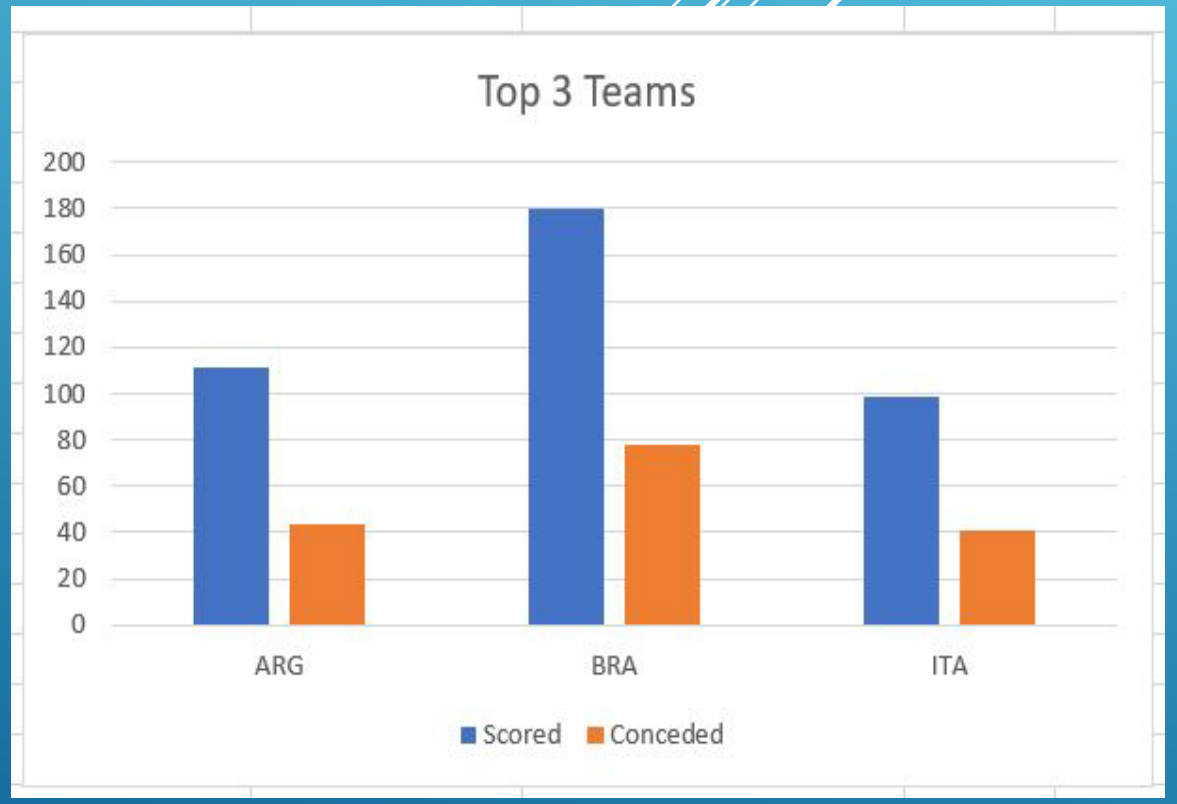
5.ANALYZING DIFFERENCE BETWEEN TOTAL GOALS SCORED AND TOTAL GOALS CONCEDED

Description:

Difference between goals scored and conceded by teams.

Row Labels	Scored	Conceded
ARG	111	44
BRA	180	78
ITA	99	41





6.HALF TIME GOALS SCORED BY THE HOME AND AWAY TEAMS

Total number of goals scored by home and away teams year by year from the year 1930 to 2014 to depict the mental confidence and usual trend of goals scored. From the trend given below we can see home team scores 604 goals in foist half and away team scores 365 goals. There is a massive goal difference leading to home team advantage.

Specific function and requirements

We have to create a pivot table to determine the difference in goals and then visualize it on graph.

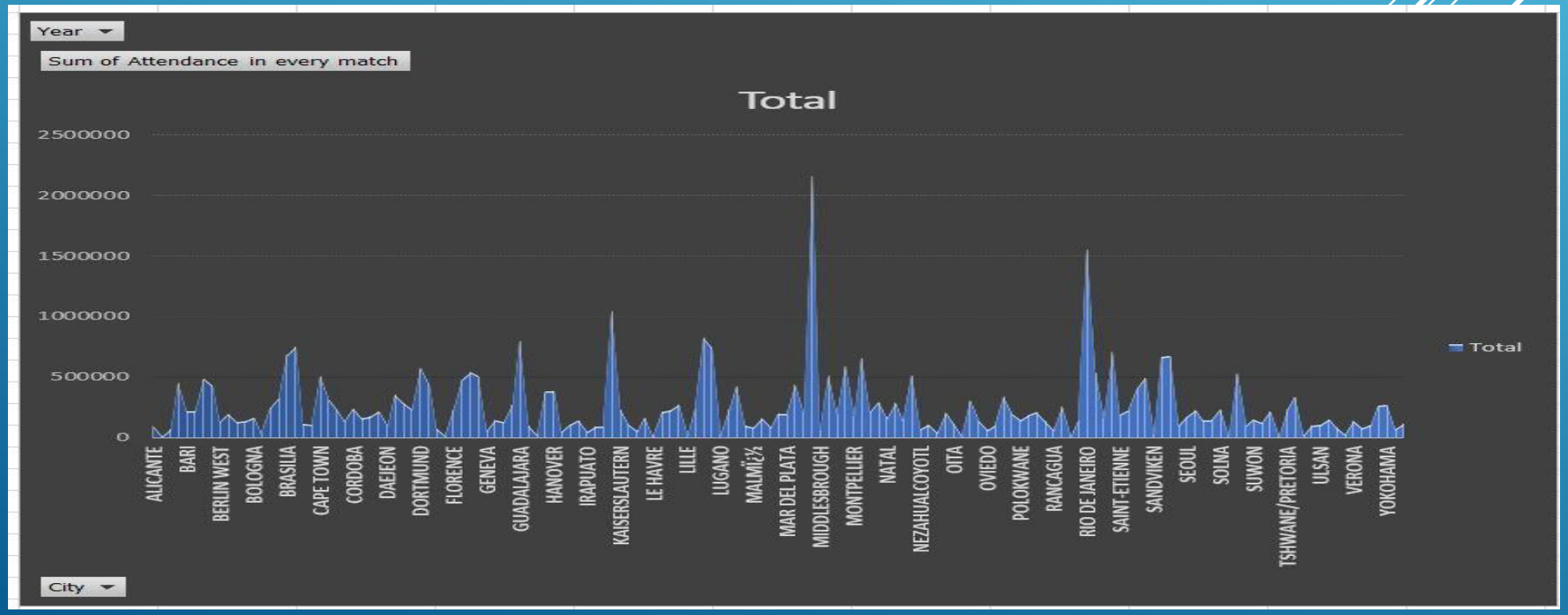
7.OVERALL ATTENDANCE FOR A PARTICULAR CITY UNTIL NOW

Description:

Determine the sum of attendance of all the people when they are home team.

Visualization:

We visualize the above results with the help of area chart created using area chart.



8.NO. OF PEOPLE CAME TO VIEW MATCH FOR PARTICULAR TEAM

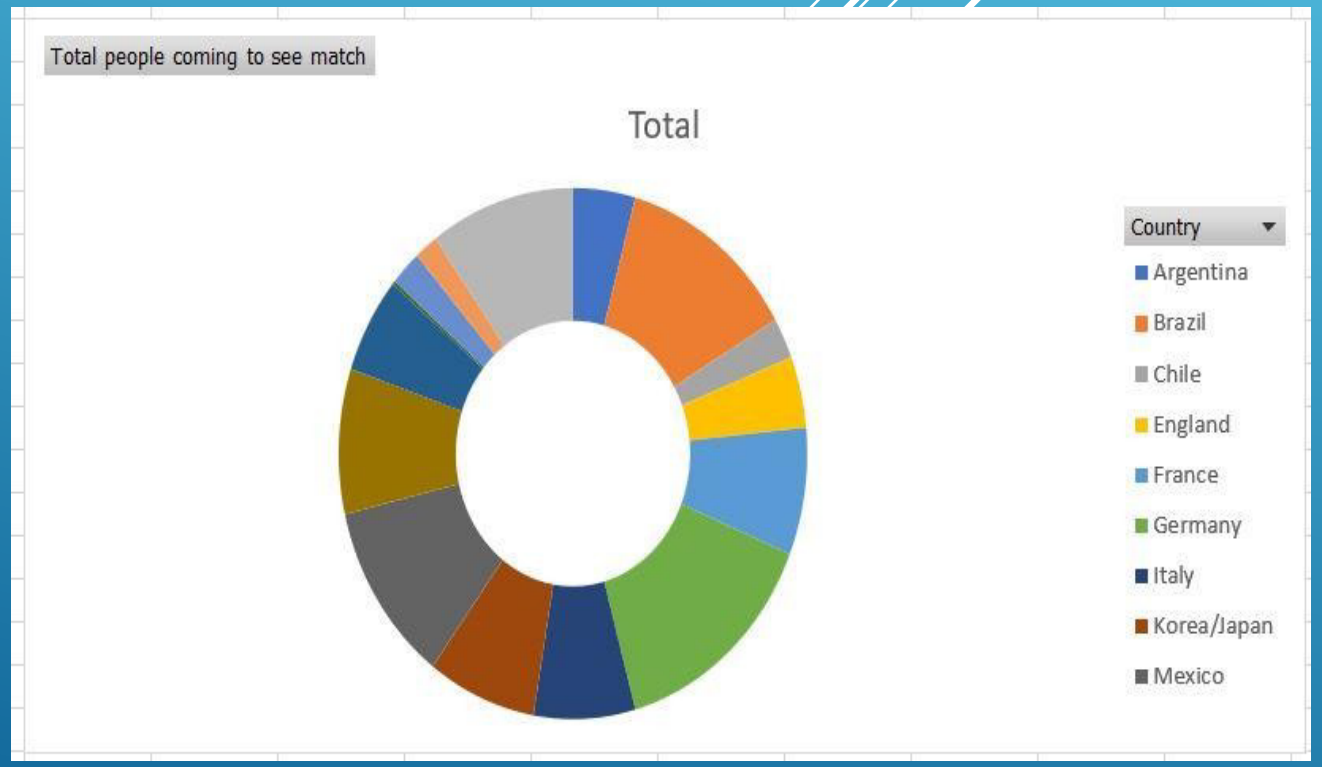
Description:

By comparing the attendance of all cities we get top 5 cities with highest audience.

Specific function and requirements:

We get cities with highest fans.here we potray top 5 countries with highest attendence and snips of data.

Country	Total people coming to see match
Argentina	1545791
Brazil	4432056
Chile	893172
England	1563135
France	2788857
Germany	5225192
Italy	2516578
Korea/Japan	2705197
Mexico	3998006
South Africa	3178856
Spain	2109723
Sweden	81981
Switzerland	768607
Uruguay	590549
USA	3587538
Grand Total	**35985238**



9.NUMBER OF PEOPLE CAME TO VIEW MATCH FOR A PARTICULAR TEAM

Most people came to see match for brazil then argentina,Italy England and so on..