**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** |  | **Student ID** |  |
| **Class** |  | **Assessor name** |  |
| **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** |  |

**Grading grid**

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

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

Table of Contents

[Introduction 5](#_Toc49426351)

[1. Overview about BI 5](#_Toc49426352)

[a. Definition of BI 5](#_Toc49426353)

[b. The effect of applying BI 5](#_Toc49426354)

[c. Real examples of BI 5](#_Toc49426355)

[2. BI techniques 6](#_Toc49426356)

[3. Tools of BI 7](#_Toc49426357)

[a. Tableau 7](#_Toc49426358)

[b. The version of tableau. 7](#_Toc49426359)

[c. Excel 7](#_Toc49426360)

[Programming tools 8](#_Toc49426361)

[Introduce to the company and the dataset 8](#_Toc49426362)

[Clean data process 9](#_Toc49426363)

[Data analysis 12](#_Toc49426364)

[Discuss the legal issues involved in exploiting user data for business intelligence. 14](#_Toc49426365)

[References 15](#_Toc49426366)

# Introduction

## 1. Overview about BI

### a. Definition of BI

Business Intelligence is "a tool to help businesses access vital business information, analyze and make quick decisions". BI gathers and processes all types of raw data in the business into graphs, vivid, easy-to-understand, easy-to-interact images, giving leaders and managers a holistic view. on the overall picture of the business, providing detailed information of all business segments in the business.

The core of a BI system is Data Warehouse and Data Mining because the data used in BI is data aggregation (Multiple sources, multiple formats, distributed history and scheduled) is a calculation. the energy of the data warehouse. At the same time, data analysis in BI is not simple analysis (Query, Filtering), but data mining techniques (Data Mining) used to classify (classify) groups (Clustering), or Project. Judgement. Therefore, BI has a very close relationship with Data Warehouse and Data Mining (Albert & Juan, 2017).

### b. The effect of applying BI

BI system helps in business and these are the priority points of BI(Albert & Juan, 2017).

* The faster report, parsing or set up a plan
* Better business decisions
* Predict the future of the business
* Retain old customers and anticipate customer functionality
* Build a business strategy
* Data quality has been improved
* Improved performance
* Improve customer satisfaction
* Increase competitive advantage
* Increase sales

### c. Real examples of BI

Examples of how BI applies to the organization:

- One of the areas of business in which BI has been most effective is the finance industry. American Express has been a pioneer of business intelligence in this sector, using the technology to develop new payment service products and market offers to customers.

BI software also helps the credit card company detect fraud more accurately and thereby protect customers whose card information may have been compromised (CCS Technology, 2018).

- a large enterprise using BI is Lowe’s, America’s second-largest home improvement store chain.

In 2007, the company started building a new data center in Texas specifically to expand on its already significant business intelligence capabilities. Like many retail chains, Lowe’s uses BI to optimize its supply chain efficiency and reduce the rate of fraudulent returns in its stores (CCS Technology, 2018).

## 2. BI techniques

There are many business intelligence techniques that businesses may use to obtain useful insights in order to make decisions. Here are some BI techniques.

**- Data Cleaning**

Data cleaning is the process of removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted. Discarded data is often unnecessary or useful when analyzing the data as it can interfere with the process or provide inaccurate results.

**- Data Visualization**

In simple terms, data visualization is a pictorial representation of a data set. For example, Text-based data graphically visualized as charts, charts, tables, infographics, graphs, and more.

**- Data Mining**

Data mining is the process of classifying and organizing large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining is an assistive method to identify unknown or unnoticed relationships in a data set.

**- Predictive Techniques**

Predictive Techniques is a BI technique that analyzes past and present data so that probabilities or trends in the future can be predicted. For example, a company that makes headphones, they may use sales data or user feedback data to be able to add features to their products or modify them as appropriate. with future trends that have been predicted.

**- Analytics**

Analytics is a technique that involves studying available data so that meaningful data can be drawn in to capture more detailed information and make more accurate decisions.

## 3. Tools of BI

### a. Tableau

Tableau is a tool to quickly, simple, and manage business analysis data for everyone, simplifies raw data in an easy to understand format.

With Tableau, even those with no programming expertise or skills can retrieve the necessary data from the data source, just intuitive manipulation by drag and drop, and display and processing of the underlying data. The recognizable quickly visualized form of data is Tableau's characteristics.

### b. The version of tableau.

Tableau is a brand, not a specific product

Tableau Desktop: Personal is the entry point for the paid development versions of the software. It allows you to keep your workbooks private, but connection and distribution options are limited.

Tableau Desktop: Professional is similar to Tableau Desktop: Personal in that it is a development version of Tableau. Both the Personal and Professional versions have all of the same development capabilities, but the Professional version provides full access to every data type and distribution channel currently available in the software.

Tableau Public is free to download, but this product actually provides development capabilities. The catch is that the workbooks have to be saved to Tableau’s public cloud, making this an unsuitable choice for proprietary business data.

### c. Excel

Both Excel and Tableau are tools for data analysis, but each method has a special data exploration approach. Yet Tableau's analysis is more effective than excellent.

# Programming tools

Python is a common high-level multilingual language used in various business and technological domains. Because of its reading ability and syntax, it is considered to be the best language for beginners. Python uses fewer line codes, making programming much simpler for the entire system. Data scientists now engage in linking network applications, Web programming, or computer automation.

Python is the best language you need to perform those tasks if you're looking for a programming language. Python is an ordinary programming language that comprises numerous modules used to analyze or view data in various tasks. The Python Business Intelligence or Software Technology Domain may use such modules as Scikits, SciPy, NumPy, or R or Disk.

Python can be easily learned and applied. You do not have previous Python knowledge, to begin with, data analyzes if you are wondering. First of all, you must change your mind. You do not need the expertise to work with data sets in the Python programming language.

# Introduce to the company and the dataset

Our dataset is collected data from the World Cup championship since 1930. Our dataset’s included fields such as: Year, Datetime, Stage, Stadium, City, Home team name, Away team name, Home/Away team goals, Win condition, Attendance, Referee, Assistant 1 and 2, Match and Round ID, Home/Away teaminitials.

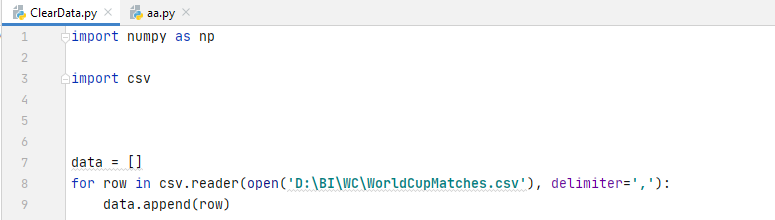
Before we start building charts and giving analysis and evaluation. Our team has some problems with data sets that need to be addressed.

* 1. In the Win Condition column, we have to change the empty data to ‘Win in 90min’
* 2. In the Referee, Assistant 1, and Assistant 2 column we have to split the column in to 2 others: Name and Home Country
* 3. We have to change the short form of the country name in 3 columns of Referee and 2 last columns Home/Away Team Initials to full form

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

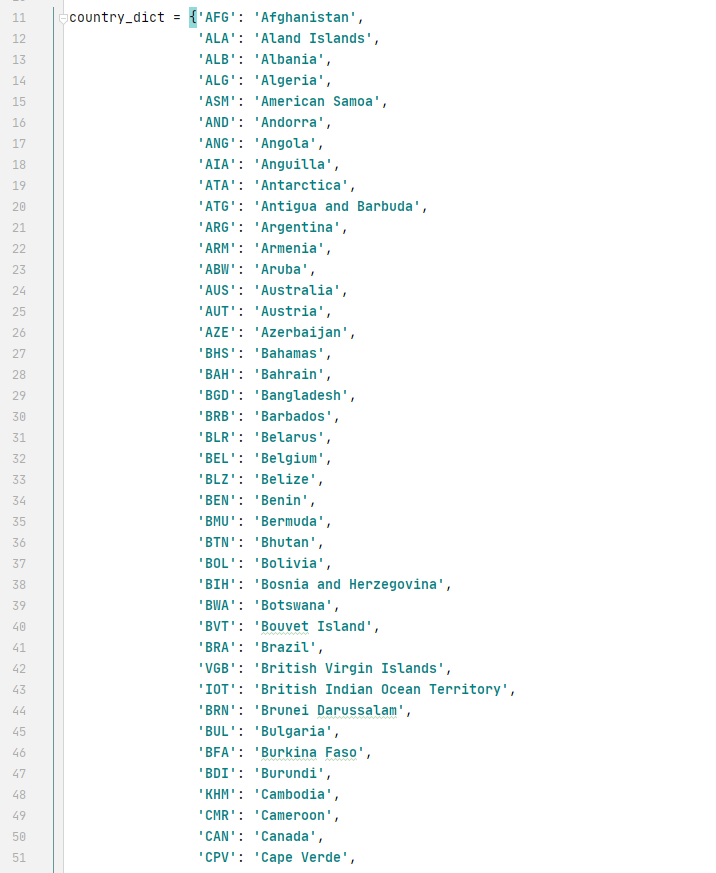
# Clean data process

Step 1: Add the required libraries and read data from csv file



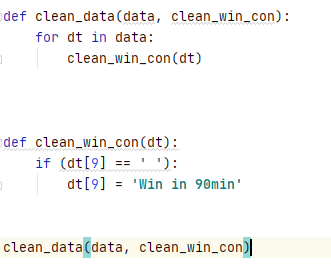
* Import library of numpy and csv
* Create list of data and set of rows which had been seen to avoid duplication rows
* Use for loop to read individual rows in csv file then insert into data list.

Step 2: Create the Library to change Country Name from short form to full form



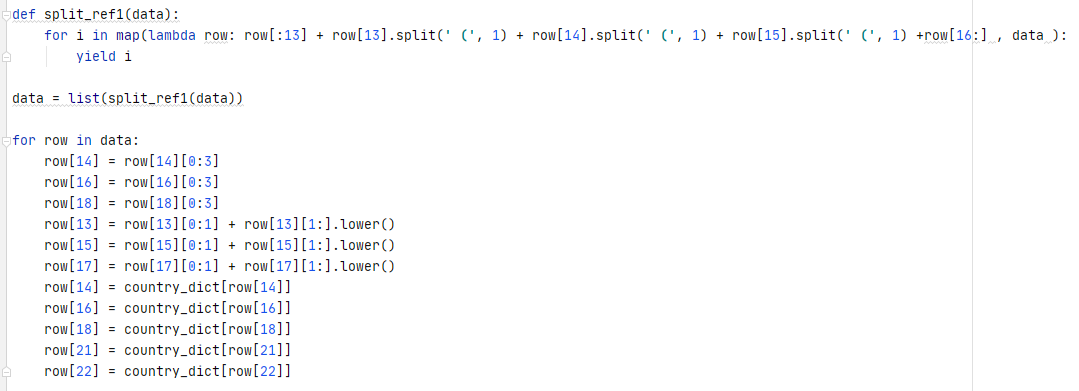
To turn country names of national team or referees home country form short to full form, I have to make a library which form is: ‘Short form’ : ’Full form’ of all countries

Step 3: Clean Win Condition Column



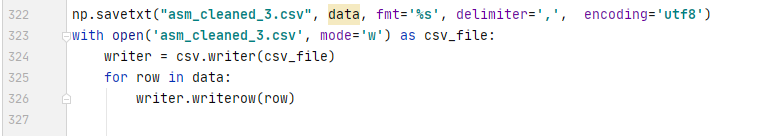
In the Win Condition column, I have to change all blank row to ‘Win in 90min’. This function will scan all the rows in Win condition column and turn all black row into ‘Win in 90min’

Step 4: Clean 3 referee columns and 2 team initials columns



Next, the split\_ref1 function will split Referee, Assistant 1 and Assistant 2 column in to 2 from the ‘(‘. However, I still have the ‘)’ in the end of all short form country name, so I have to erase it by take the data from 0 to 3 in each column. After that, I have to turn all name of Referee and Assistant to lowercase except the first character. Finally, I only have to change short form name to full form by using the library I created above.

Step 5: Write the project to .csv file

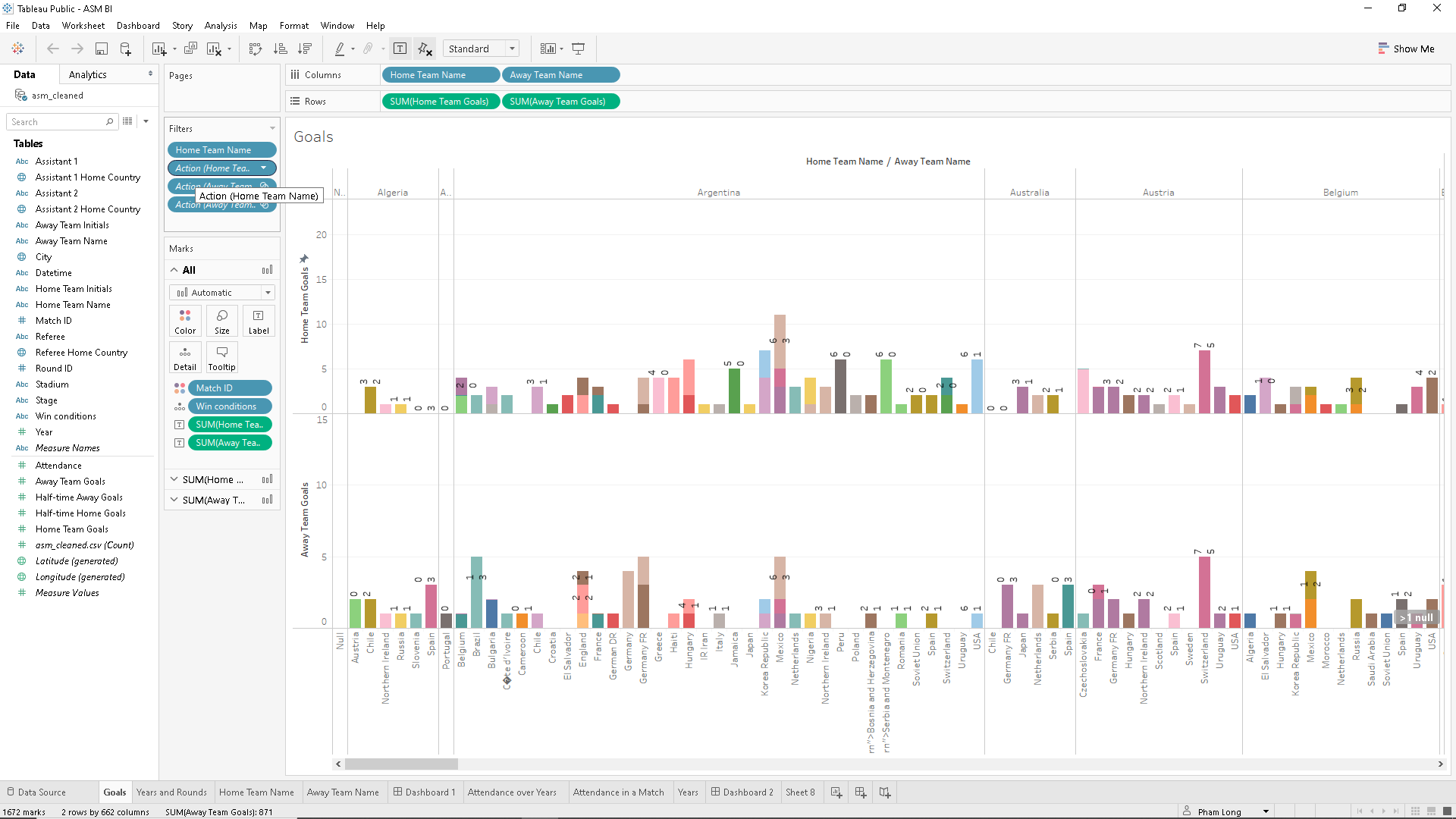


This block will write the data variable in to a new csv file which is called asm\_cleaned\_3.csv.

# Data analysis

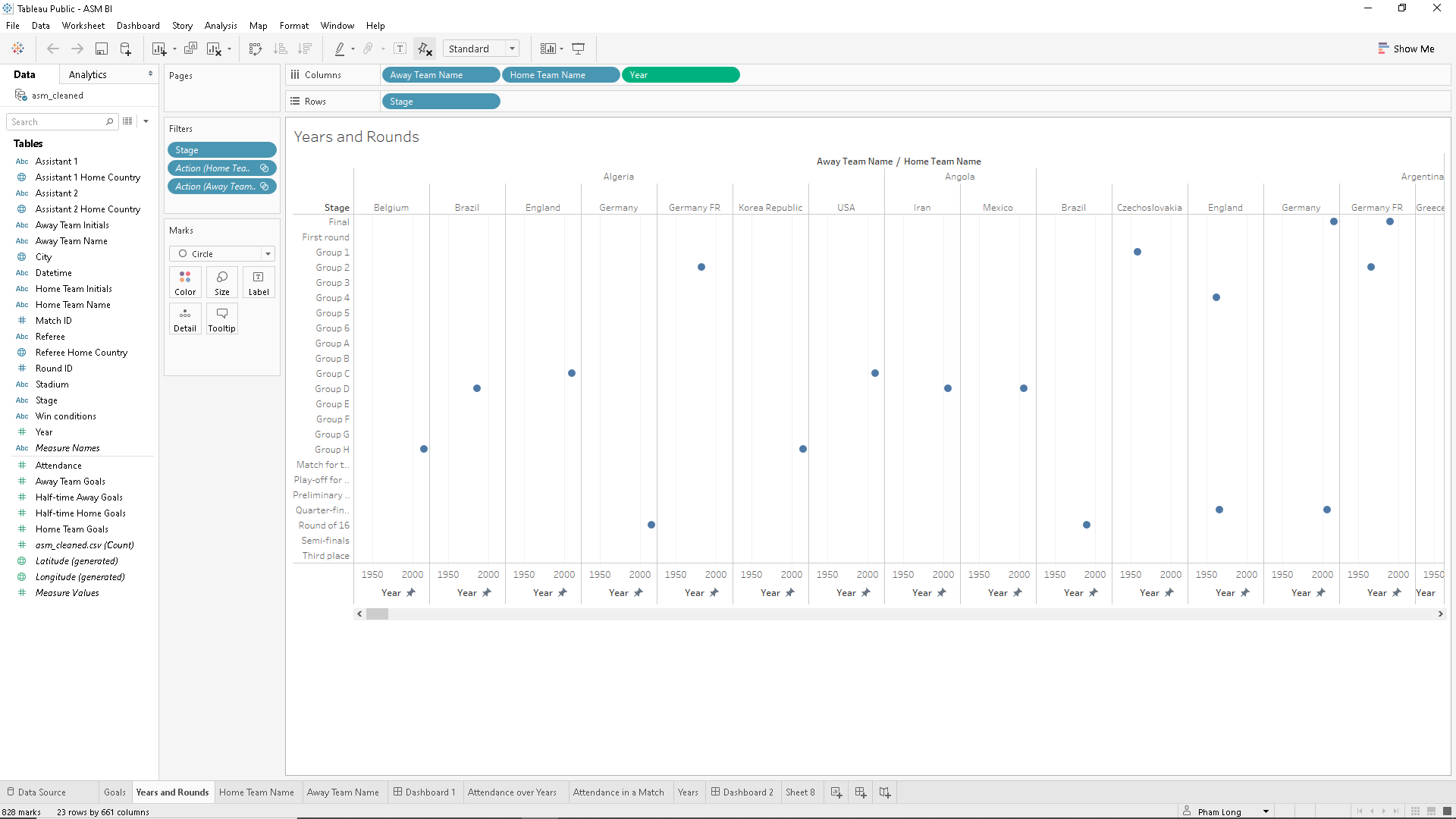
After representing the cleaning data step, now the clean data will be analyzed using tableau.

Chart 1: The results of matches from the pass

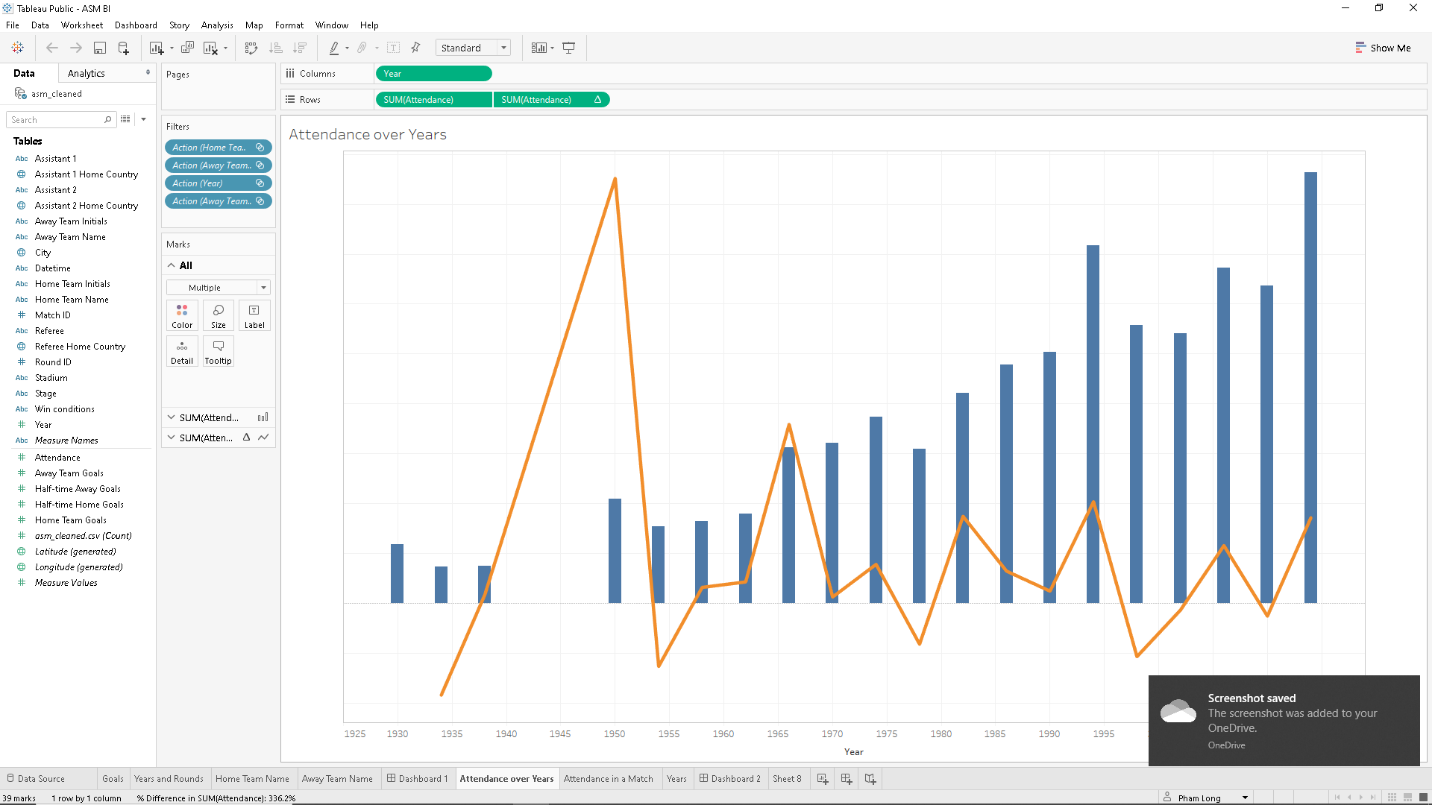


This chart will show all results of World Cup matches from 1930, which included 2 team name and the result of matches between those 2. In this chart, Home Team will be the team that wear home shirt, and Away team will be the team that wear away shirt.

Chart 2: Time matches happened



In this chart, we can see the time matches happened, which teams attended in that match, those 2 teams meet in which round.

Chart 3: 

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

Privacy is a big issue when mining user data. So, this section discusses Privacy.

Privacy is now widely recognized and protected by national and international laws. Each individual should have the right to know that his or her personal information is collected and maintained by whom, in the where and for what purpose. Additionally, each individual has the right to request correction or deletion of his / her personal information if it is being stored inaccurately, or is illegally collected or stored. Privacy, however, is not an absolute right, it can be restricted under certain circumstances for the common good of society.

So be careful in exploiting user data. There are a few guidelines that help us avoid a privacy breach:

* **Limited collection**: The collection of personal data and any such data should be done by legal and fair means, and where appropriate, with knowledge or consent. of data object.
* **Identified collection purposes**: Personal data collection purposes must be defined no later than the time of data collection and then limited use in the performance of those purposes; It would be inappropriate to have such collection changed in purpose.
* **Limited Use**: Personal data is not disclosed, made available for use for purposes other than the original ones, except with the consent of the data owner or as required by law. concentration
* **Data security**: Personal data must be protected with reasonable security safeguards against such risks as loss or unauthorized access, use, destruction, modification, modification, or moderation. data disclosure

# References

CCS Technology. (2018)Real-world examples of business intelligence Available from: [https://www.ccstechnologygroup.com/](https://www.ccstechnologygroup.com/?fbclid=IwAR31k-o3Mly7feAae1wLr836kjq4Z-V4-zfuW7fmwOAChs6bLxHFqCjPi1g)

[Accessed 20th Aug 2020].

Albert Nogués, Juan Valladares. (2017) Business Intelligence Tools for Small Companies. Spain, apress Publishing.