Introduction

I participated in the data cleaning challenge on the twitter space. This was an opportunity given to Data Analyst to improve on their data cleaning skill. My previous Projects Data were either clean or semi clean datas, that is the main reason why I participated in the #datacleanining challenge as this was a messy dataset, this was to improve on my cleaning skill and learn from others.

About the Data

The FIFA 21 dataset was gotten from the Kaggle website <https://www.kaggle.com/datasets/yagunnersya/fifa-21-messy-raw-dataset-for-cleaning-exploring>, It was a raw and messy csv data which contained 18979 rows and 77 columns. The dataset contained informations about football players which included their IDs, Names, Height, Weight, best positions, Wages, Value and so on

DATA TRANSFORMATION

Before the data was transform, I noticed the data has special characters (), I then changed the file Origin from Western European to Unicode UTF8 as seen below. Thereafter proceeded to transform data.

DATA OBSERVATION AFTER IMPORTING

1. No empty spaces
2. White spaces
3. No Initial Errors
4. Wrong data types
5. Irrelevant columns
6. Incorrect values

DATA CLEANING PROCESS

Power Query Editor was used for the data cleaning. After importing the data, the whitespaces were removed using the Trim function.

Graphical user interface

Description automatically generated with low confidence

Graphical user interface, text, application

Description automatically generated

I started the data cleaning process from the first column to avoid omitting any column.

**ID Column**

I changed the data type from whole number to text as IDs are unique and will not be used for calculation

Graphical user interface, application, table

Description automatically generated with medium confidence

**Name Column**

The names were correct except 1 which had special character as name initial, I used replace value function to change it to the correct initial

Table

Description automatically generated

Table

Description automatically generated

**Overall and Potential Rating Columns**

Both columns data types were changed to percentage (%) as required by the data dictionary and the names were written out in full for better understanding

Table

Description automatically generated Table

Description automatically generated

Contract Column

On the raw data there were different contract which were ungrouped. I created a new column(Agreement) and grouped the contracts accordingly. I also separated the start and end dates. The Contract duration was also calculated

Text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Height Column

The measurement for the height in the raw data were in cm, feet and inches, they needed to be in the same measurement which is cm. I first converted the feet to inches (feet\*12 = Inches), added both inches together before converting to cm (0.393701\*inch = cm)

A picture containing graphical user interface

Description automatically generated Table

Description automatically generated

**Weight Column**

The measurements for weight were inially in kg and lbs. I converted kg to lbs to have a uniform measurement.

Graphical user interface, text, application

Description automatically generated with medium confidence Graphical user interface, text, application, email

Description automatically generated

**Value, Wages and Release Clause Columns**

The M and K were replaced with 1000000 and 1000 respectively, € was removed and data type was changed to $

Table

Description automatically generated

Table

Description automatically generated

**W/F, SM and IR Columns**

The special character (★) was removed, the abbreviated headings were written out in full for better understanding. The data type was changed from text to whole number to aid calculation during visualisation.

Table

Description automatically generated

Table

Description automatically generated

Hits

This column had some values which ended in K, the were replaced by 1000 using multiplication

Table

Description automatically generated with medium confidenceTable

Description automatically generated with medium confidence