Data Mining Project document for project Prediction of 5-Year Career Longevity for NBA Rookies.

Spring Semester 2023

**CS334**

**Name:** Louay Mustafa.

**ID:** 184249.

* **Title**: Prediction of 5-Year Career Longevity for NBA Rookies.
* **Data description:** A description of the data in detail under the following subheadings:
  + **The problem domain**

The dataset includes 21 attributes and 1341 values. Most of the values are numeric. The class attribute has the values 0 and 1 with **0 being ‘No’ and 1 being ‘Yes’** This a dataset consisting of NBA rookies and our task is to predict if they will be playing in the NBA after 5 years.

* + **The source of the data**

<https://data.world/exercises/logistic-regression-exercise-1>

The dataset is available on data.world and is available as a classification exercise.

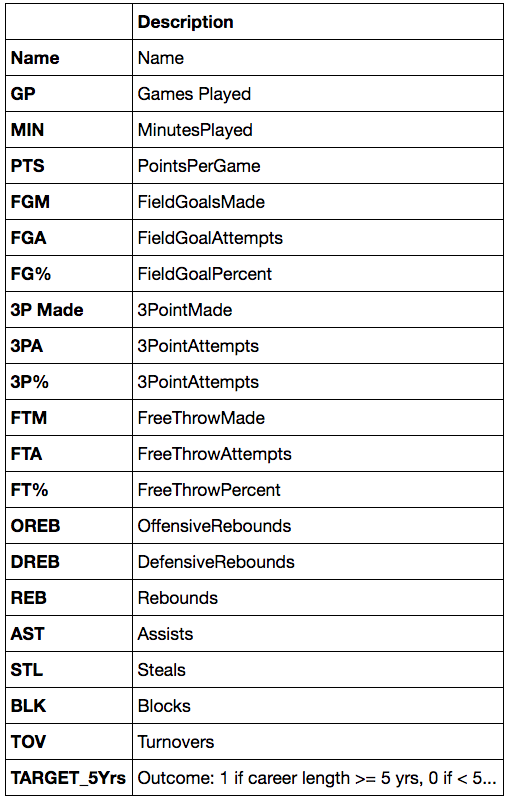
* + **The intended use of the data**

The purpose of the data is to help NBA coaches select a player that can last in the league for a long term.

* + **The attribute types of the data**

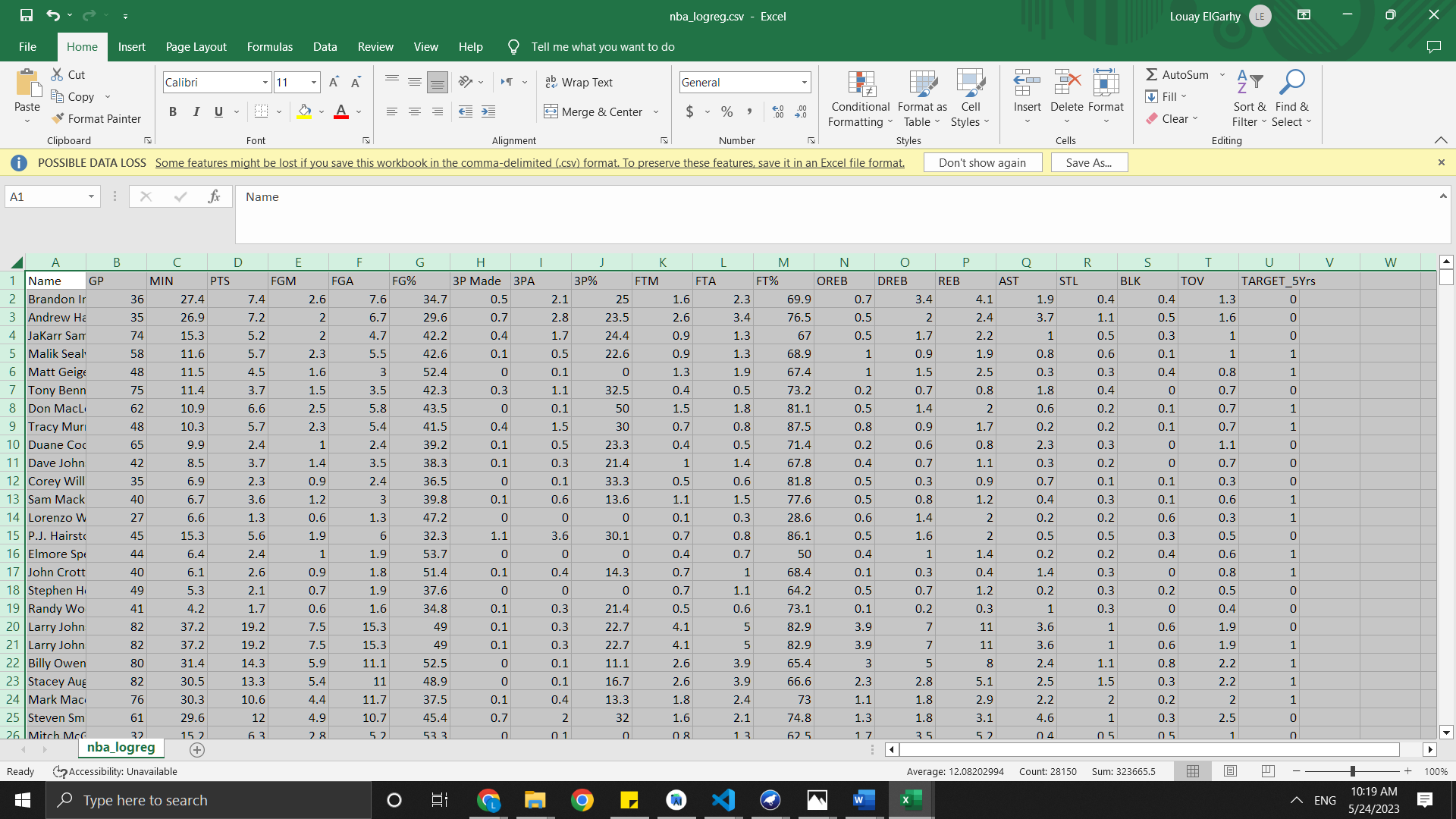
Nominal and numeric

**Table**: Details of attribute names

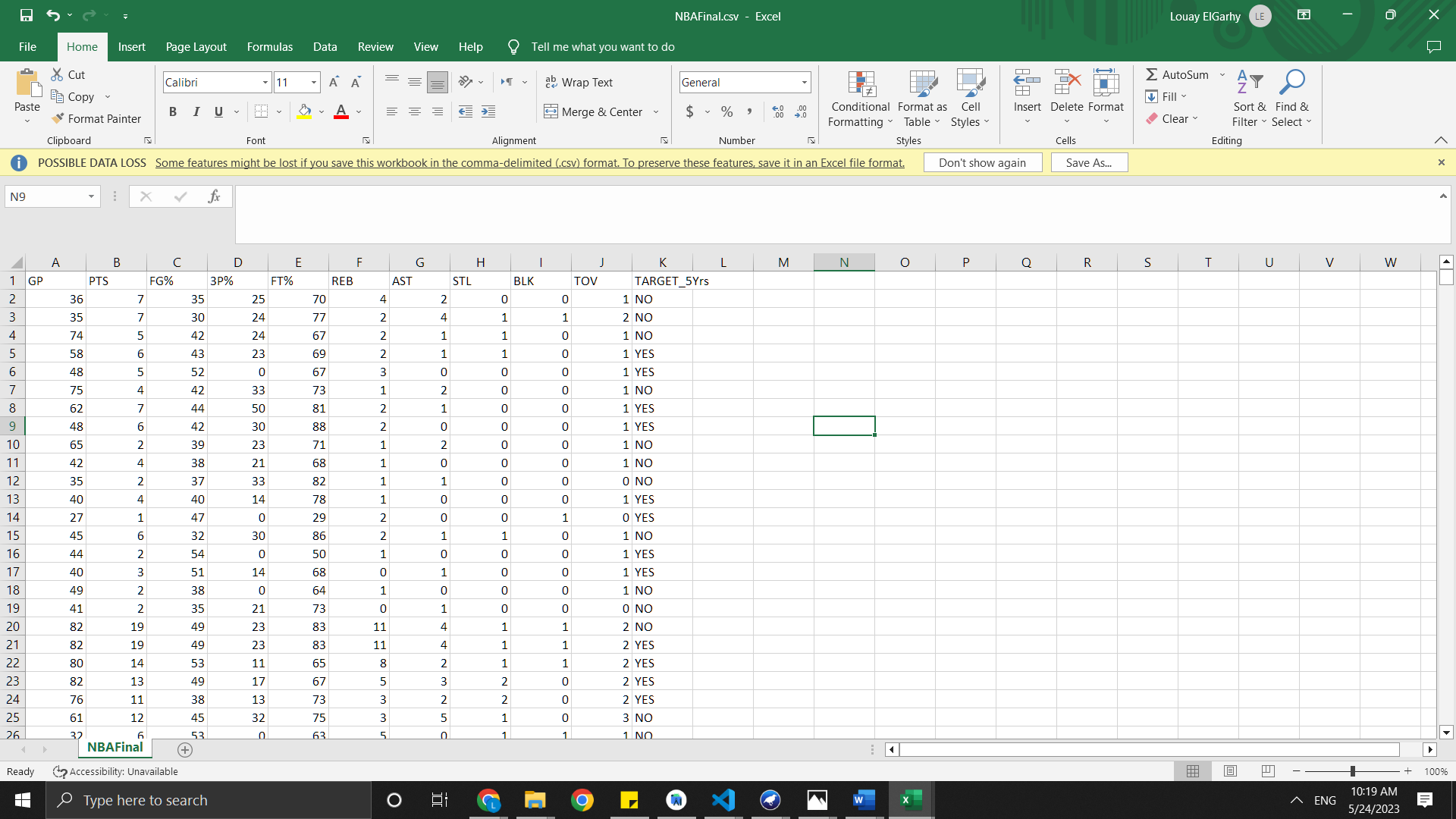


* **Objective**: To Predict 5-Year Career Longevity for NBA Rookies.

**Names of players** was removed as it had many unique values which would cause overfitting and had no value whatsoever. **MIN** (minutes played) was removed because we are already using games played attribute. **FGM**(Field goals made),**FGA**(field goal attempts),**3P Made** (3 points made),**3PA**(3 point attempts),**FTM**(free throws made),**FTA**(free throw attempts),**OREB**(offensive rebounds),**DREB**(defensive rebounds) are all removed since we already have attributes like **FG**(field goal percent),**3P**(3PointPercent) , **FT**(Free throw %),**REB**(rebounds) that provide the same use and removes redundancy. All the values are numeric and need to be discretized.



These are the attributes of the dataset after discretization on a **Microsoft Excel**. However, these values cannot be in fractions and hence it actually required to manually convert them to whole numbers.



These are the attributes of the dataset after converting the discretized values from fractions into whole numbers. The class attribute is **TARGET\_5Yrs** with the 0 being ‘**No’** and 1 being ‘**Yes**’. The dataset is now ready for classification, Clustering and association as well.