<https://www.kaggle.com/datasets/ilyadziamidovich/league-of-legends-world-championship-2022?resource=download&select=wc_champions.csv>

Did Aatrox help teams win more in LoL Worlds.

In this data analysis we will be looking at whether or not Aatrox was picked and if so did the team who picked him win. Other factors that can contribute to this data is the ban rate and what side they were picked on either being red or blue. Blue is first pick side with also first ban while red has two first picks. The problem we are looking at is to see if Aatrox is overpowered. If he is then this would require a balancing of the champion, because one champion's presence shouldn’t allow a team to have a win. This game is based on map control and teamwork.

Logistic Regression/Sigmoidal Curve

When the Team has the champion selected did they win

Champion Selection Win

Column 1 Aatrox Selected by team yes no

Column 2 Team won the match yes no

Getting Data to be analyzable data

1. Problem Definition DONE
   1. Go Back to Trello and Flesh out
   2. Write the Problem Definition
      1. Write a page of the Data Analysis Workflow
      2. Section 1 problem Definition and explain what I am doing with the data
         1. Developing a model that will allow us to predict if Aatrox is overpowered and causing teams to win with him. Does the selection of the champ change the outcome of the game without even playing the experience. Selection of Aatrox vs Match Outcome.
2. Data Acquisition/Data Exploration
   1. Find and explore my data
      1. Official Title of the Data
      2. Description of the data
      3. Url linking to the data
      4. Want some statement of the currency of the data
      5. Pull the data found and make notes and explanation
         1. Column 1 Aatrox Selected by winning team yes no
         2. Column 2 Team won the match yes no
      6. See If I can find a linear regression in the data I have already.
3. Data Cleansing and Preparation
   * 1. Missing Data can be dropped or imputed
     2. Validate and Format the data to apporita types
     3. Reasonability/Outliers
        1. Example Age of player instead of 25 someone typed 205 so you need to look and check to see if there are outliers where under a certain valid value and over a valid value
        2. Outlier Analysis of how far the data point is far away from the mean
        3. Scaling is a very model specific thing you want to read the details of the library you are using to see how sensitive the scale is
           1. Look at Keggal for data example to practice clean data
     4. Making Sure the data is in the right format
        1. Making sure the data variable are able to be read
        2. excel, json, csv
     5. Making sure the data is there
     6. Removing duplicate value
        1. =IF(OR (Value=Below, Value=Above), "DUPILICATE FOUND","")
           1. This will count the actual duplicate and not the duplicate pairs
4. Model Development and Analysis
   1. Model Selection
5. Presentation
   1. Summary/overview paragraph
   2. LOL overview and Gameplay Summary
   3. Talk About the Analysis and what want to be done
   4. Data collection that was done and did
   5. Summary of the data results and found

Make a Virtual Environment in VSC

Read Upon

VENV ships with python

PIPENV virtual environment

Need to instal PIP globally

Also install PIPReqs

After Running Look at the requirements.text file and see whats in there

Pip Freeze it will print out the list of all package in the current environment

Talk about how it was easier and simpiton and elegant data that was there and can give a trivial response

Find Out Linear Regression Variable

This is ultimately a question of simple probability and now going forward of a less trivial nature

**Be able to teach how Simple Linear Regression and Logistic Regression Binary**

**What is it create a logistic regression of creating a independent 1 into a dependent 1**

**Be able to demonstrate on what is Logistic Regression and what does it mean to be coded into python**

**Linear Regression/Logistic Regression Binary/complex**

**Slope and Intercept**

**slope of the logistic regression curve is steepest at this halfway point. where logit(x) = log(x/(1−x)) is a function mapping the range (0, 1) to the range (−∞, ∞)**

**the expected value of Y when all X=0.**

**y=mx+b to assign a function to make a logistic regression**

**What is the r value what is the correlation of the charts/strength of correlation**

**Put a confidence interval in all linear repression 90%**

**Fix the x and y axis to have the same value**

**Look at the additional Matplot to have hard set values**

**Put a text box for the r collation number**

**n= is the size of the population**

**Why are the Top lane and jungle lane goin down and why there is a descending correlation**

**Businesses are Profit motivated.**

**Explain on the next step on how would you**

**Does having more experiences in the Pro league give better gold per average.**

**(Will you have a higher gold per minute score) with the (higher amount of years of experience), you have played pro league.**

**In professional LOL play, is there a high correlation between years of experience and gold per minute?**

**In professional LOL play, what is the correlation between years of experience and gold per minute?**

**If you populate more points from the different things You want to wiggle one variable at a time.**

**Five Things need for data analyzed**

**Tell Meaning Full story**

**Title the chart**

**Label your axis**

**Cite your source and data population**

**If Linear regression you need to include the R value**

**Do an ethical job of explaining the attributes**

**First cut**

**Reads in the data array set of 1/0**

**In python first thing I will have to do is reshape the array to make sure that they want a 1 by x array**

**Use appropriate library in order to**

**Use functionality of the data to split between training and test sets**

**Use the test and training sets to fit and train the model**

**Should have a predictive model**

Stretch Goal

-Can be refactored to be able to generalize the code to digest all the data to get every champion instead of just the one champion

Look Up how to write a snip it of python code to look at two table values

Look more into the math Linear Regresssion

Linear Regression with continuous

What Percentage of teams selected the campion vs Teams that didn’t

**To what extent does an independent variable influences the dependent outcome or state**

**Correlation is not Causality**

Logistic Regression

For the Winning Team did the champion selected win the match

Logistic Regression Read UP on For next session To compared to Linear Regression

For every game played in worlds, was the champion pick being present, and did the team win by that team.

Team Ranking

Number of years playing

Which players in each match and what team won each match, for all the players vs years of experience.

Calculate average years of experience vs win rate

<https://www.ibm.com/topics/logistic-regression#:~:text=Resources-,What%20is%20logistic%20regression%3F,given%20dataset%20of%20independent%20variables>.

<https://www.ibm.com/topics/logistic-regression#:~:text=Resources-,What%20is%20logistic%20regression%3F,given%20dataset%20of%20independent%20variables>.