



INTRODUCTION TO DATA ANALYSIS

# CSS 328 Project Presentation

Made by Georgiy, Dias, Nurdaulet and Danial

# introduction

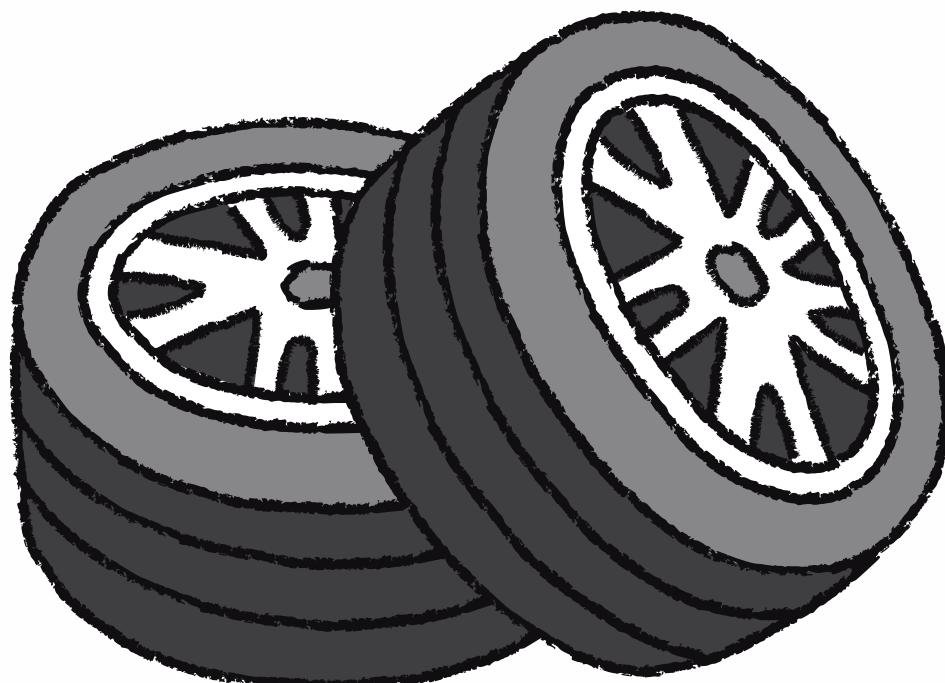
With the onset of winter, the demand for the purchase of tires increases, so we decided to take tire research as the basis for the project. We decided to do a tire research on a marketplace such as Fortebank Market since it provides large amounts of data and provides all the requirements (rows and columns). We will see various visualization graphs of the collected data and also, using machine learning models, we will see how some attributes of the tire affect its size, price, etc.



## Data Collection



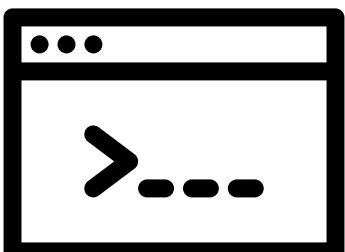
We decided to do a tire research on a marketplace such as Fortebank Market, since it provides large amounts of data and provides all the requirements.



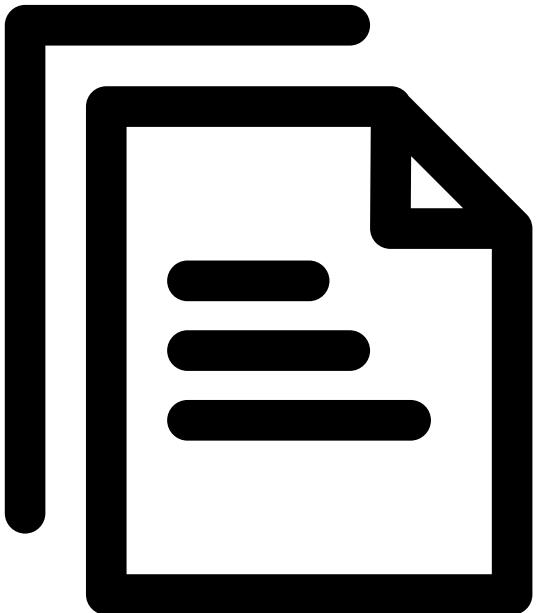
# Data Collection

Two Python files were created to collect detailed tire data from the forteMarket website, utilizing Selenium, BeautifulSoup, and Pandas.

The first file generates unique tire links, while the second file parses detailed information using Selenium and XPath locators, storing the data in a dataframe and saving it to .csv and .xlsx files in the DataSets folder.



1



tires\_link\_generator.py

2

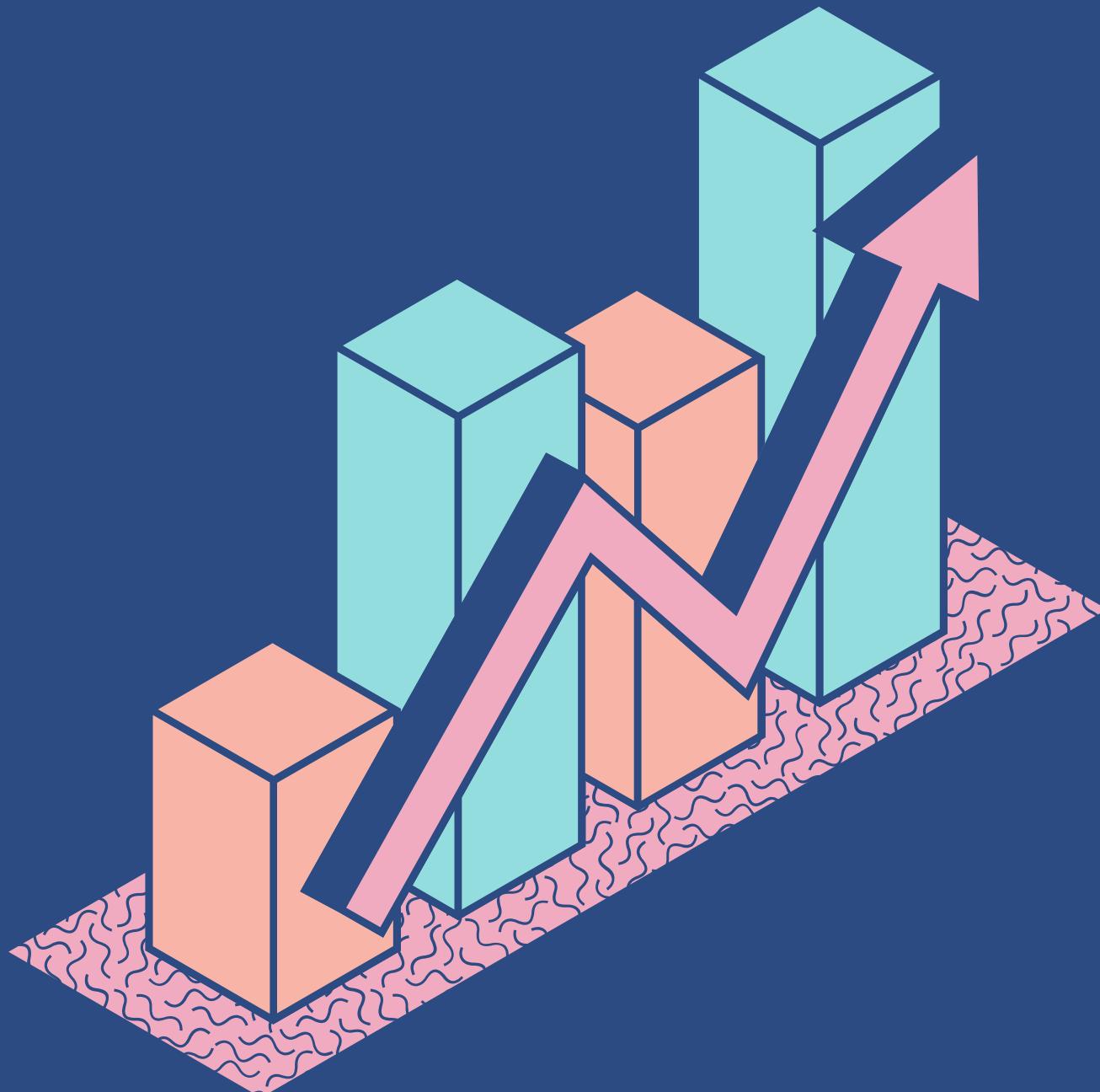


forte\_tires\_parser.py

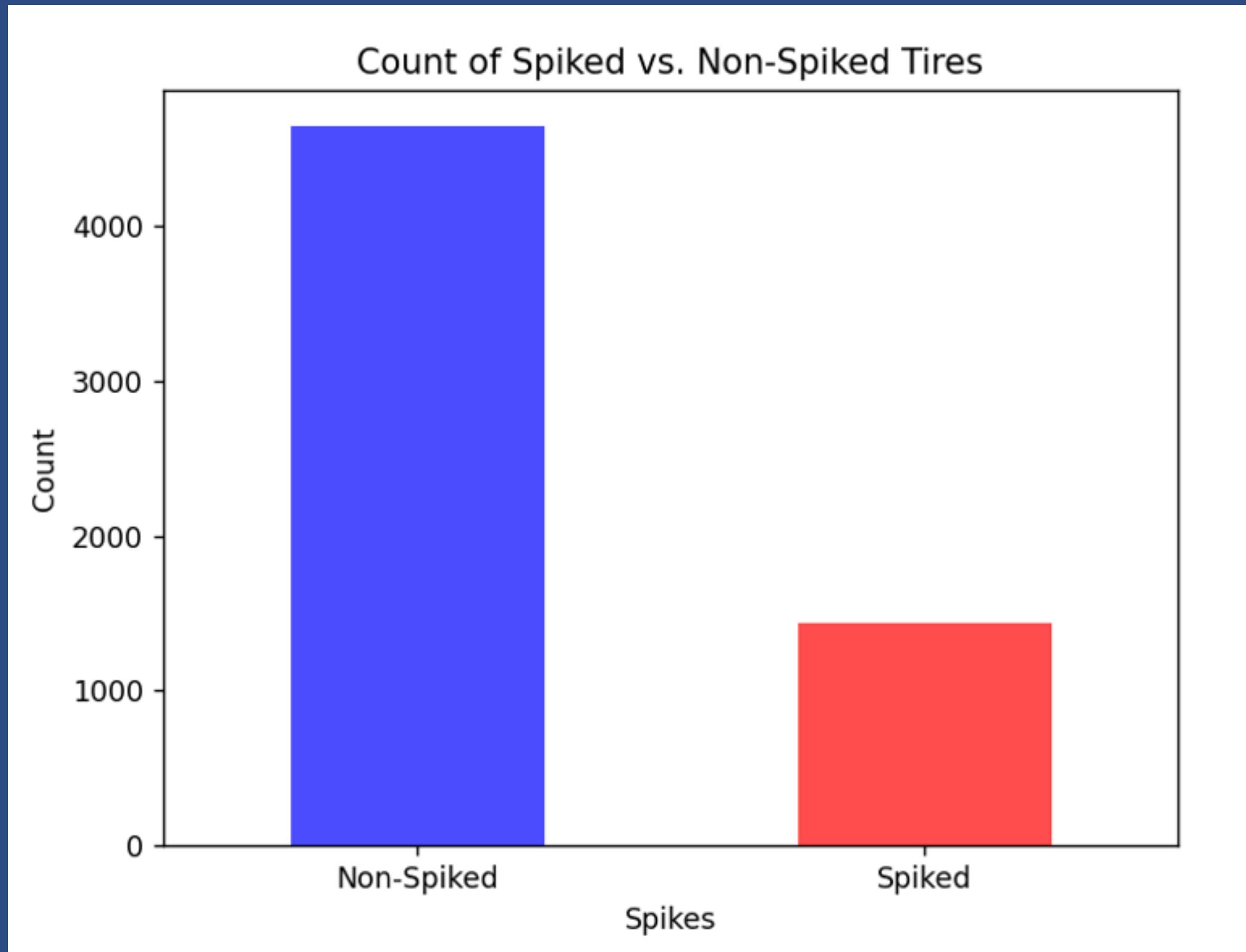
# Data Visualisation

In the visualization part, we used 5 types of graphs, such as: diagram, histogram, pie chart, plot and scatterplot.

In this presentation we will show and cover some of them.



# Diagram



In this diagram we can see statistics about how many tires are spiked and not. We clearly can see that there's more non-spiked ones, because spikes can be effective in providing traction on icy roads, that is common occurrence on the winter, but in the any other part of year it is more right to use non-spiked one.

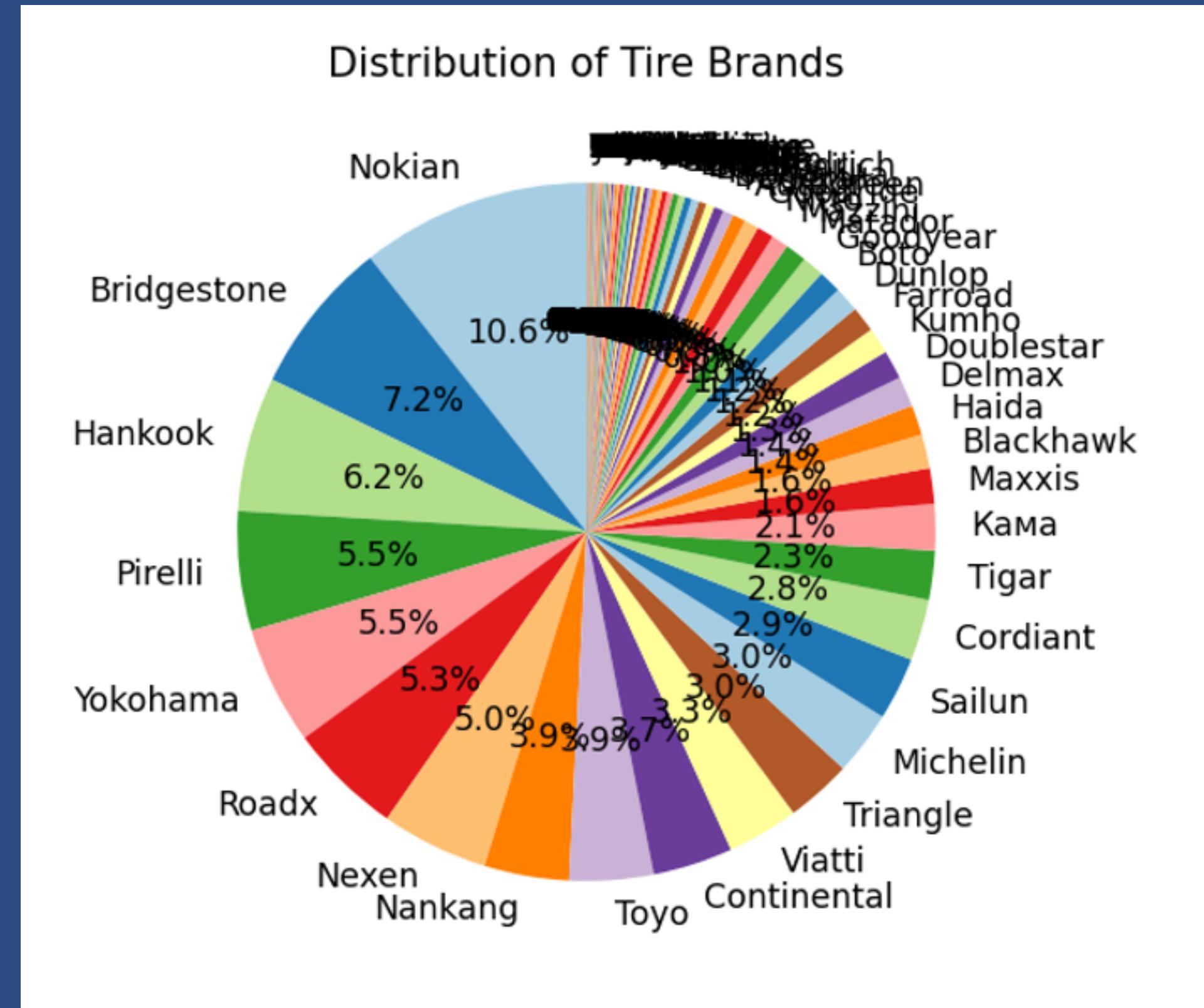


This pie chart shows us the frequency of appearance of certain brands, and their rating according to the percentage of appearance.

To construct this and previous graphics, we used python libraries such as pandas and matplotlib.pyplot

There was used very useful method `value_count()` that helped us to count values in corresponding columns.

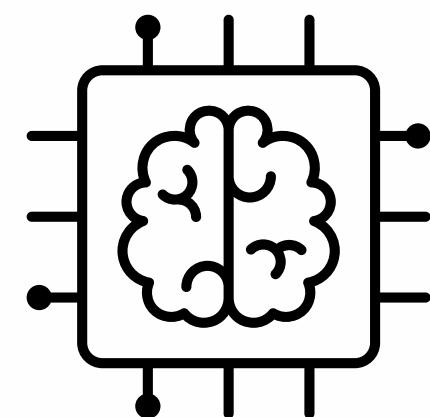
# Pie chart





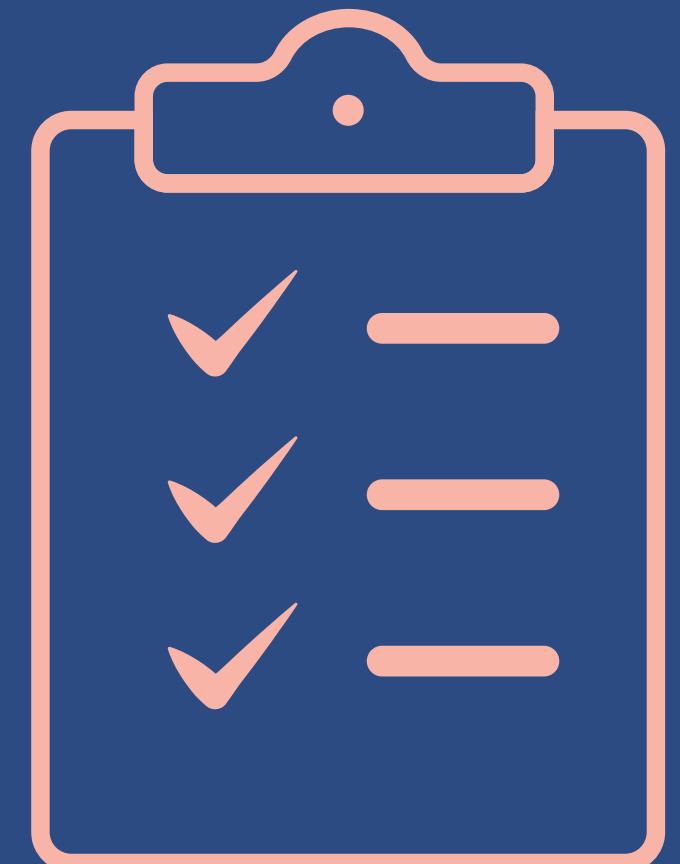
# Data Analytics

As process of data analytics and initialising ML models is complicated process, there we'll just cover steps that we did to completely analyze and modify our data.



- Data Preprocessing
- Handling Missing Values
- Creating Dummy Variables
- Standardizing 'Spikes' Column
- Handling 'Diameter', 'High', and 'Width' Columns
- Saving Preprocessed Data
- Feature and Target Variable Definition for Linear Regression
- Train-Test Split
- Model Initialization and Training
- Model Evaluation
- Classifier models using
- Model Initialization and Training
- Iterating Through Classifiers
- Evaluation Metrics
- Identifying the Best Model

**For more information, look at the documentation!**



Thank  
you!