

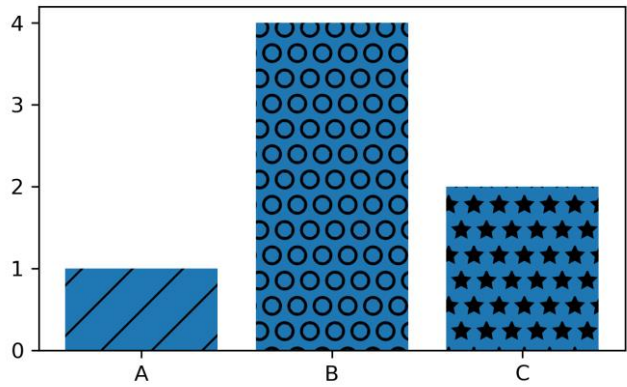
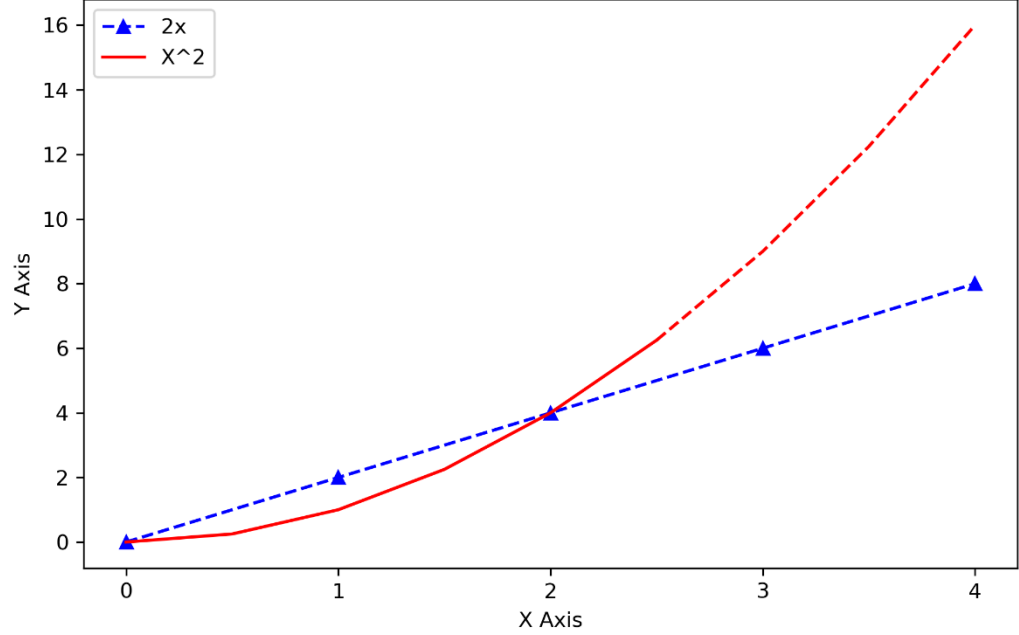
Project Name: Visualization With Matplotlib

Table of Contents

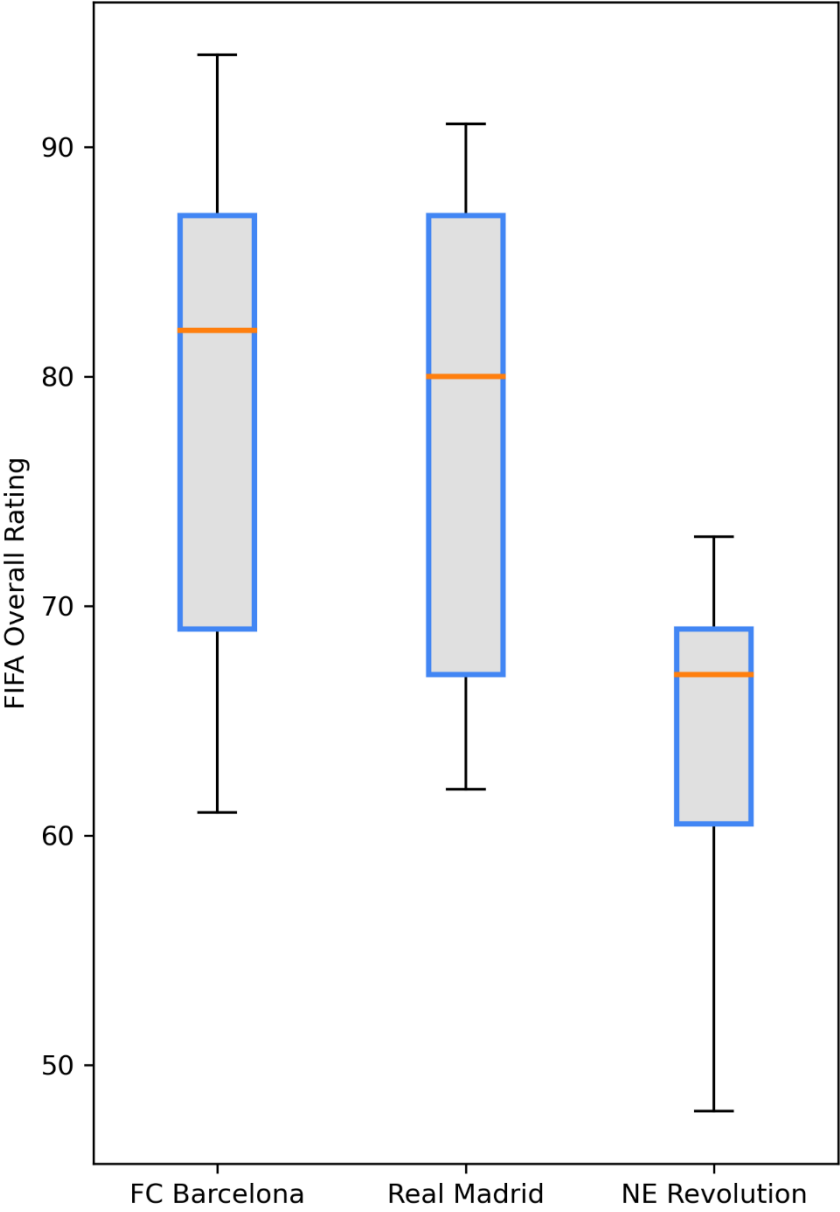
- Demo
- Overview
- Motivation
- Technical Aspect
- Installation
- Run/How to Use/Steps
- Directory Tree/Structure of Project
- To Do/Future Scope
- Technologies Used/System Requirements/Tech Stack
- Credits

Demo

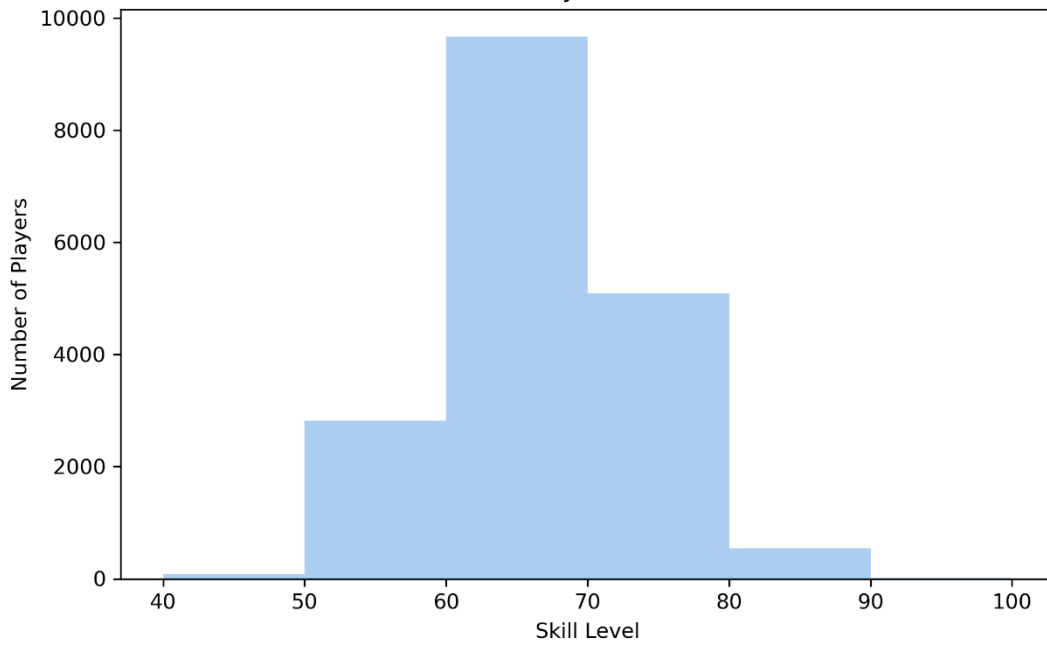
Our Graph for Visualization!



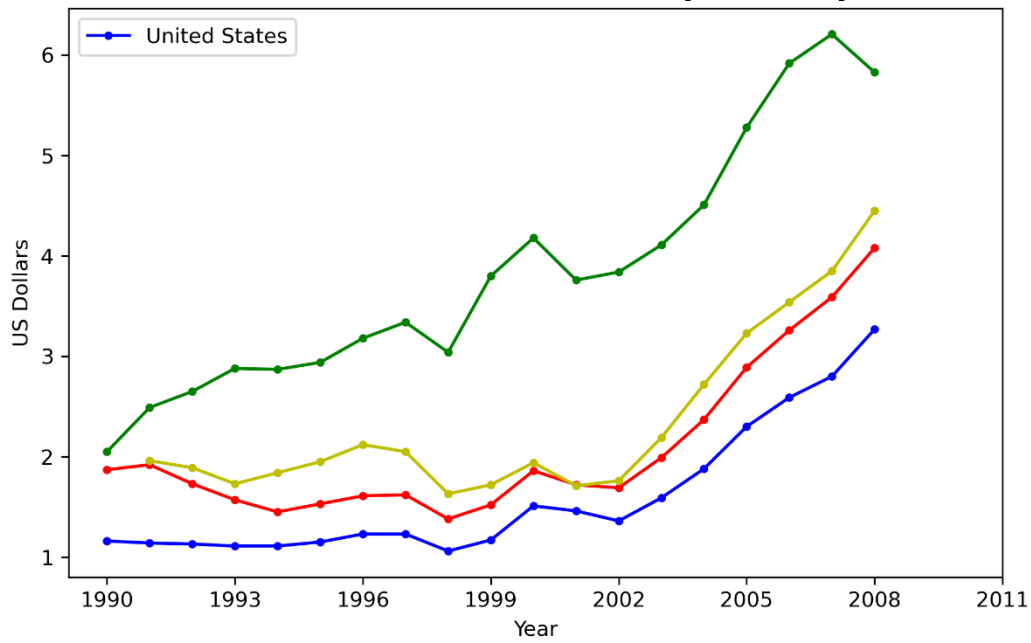
Professional Soccer Team Comparison



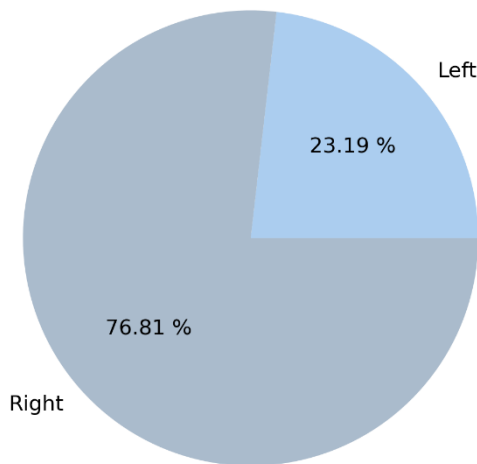
Distribution of Player Skills in FIFA 2018



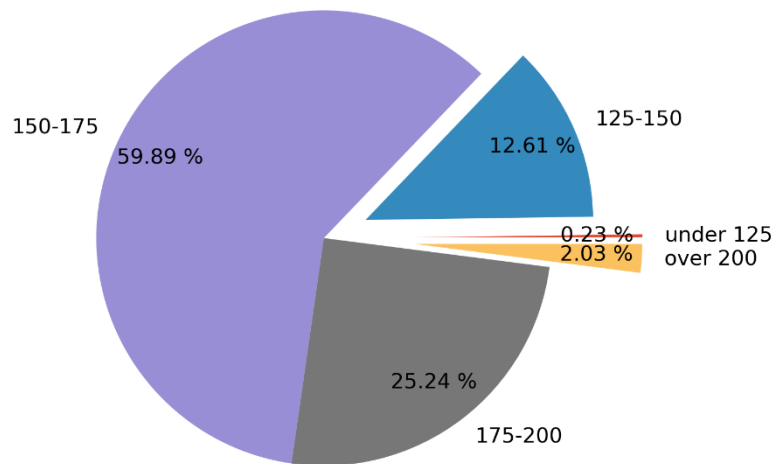
Gas Prices over Time (in USD)



Foot Preference of FIFA Players



Weight of Professional Soccer Players (lbs)



Overview

This is an Exploratory Data Analysis with Visualization using matplotlib library

EDA is essentially a type of storytelling for statisticians.

It allows us to uncover patterns and insights, often with visual methods, within data.

EDA is often the first step of the data modelling process.

This repository contains the code for EDA along with Visualization using python's various libraries.

It used numpy, pandas and matplotlib libraries.

These libraries help to perform individually one particular functionality.

Numpy is used for working with arrays. It stands for Numerical Python.

Pandas objects rely heavily on Numpy objects.

Matplotlib is a plotting library.

The purpose of creating this repository is to gain insights into EDA with Visualization.

These python libraries raised knowledge in discovering these libraries with practical use of it.

It leads to growth in my ML repository.

These above screenshots will help you to understand flow of output.

Motivation

The reason behind building this is, to maximize I as analyst's insight into a data set and into the underlying structure of a data set, while providing all of the specific items that an analyst would want to extract. It is a way of visualizing, summarizing and interpreting the information that is hidden in rows and column format. EDA is understanding the data sets by summarizing their main characteristics often plotting them visually. This step is very important for me especially when I arrive at modelling the data in order to apply Machine learning. Another motive is, as a master's student I have learnt Data Mining Subject which has somewhere led me to also learn about EDA and Visualization. Matplotlib is one of the basic libraries which everyone in DS field should know therefore polishing basics is always needed. Data visualization is the discipline of trying to understand data by placing it in a visual context so that patterns, trends and

correlations that might not otherwise be detected can be exposed. Visualization through visual imagery has been an effective way to communicate both abstract and concrete ideas since the dawn of humanity. A good visualization tells a story, removing the noise from data and highlighting the useful information. Effective data visualization is a delicate balancing act between form and function. Even statistically, it is said that child from 0-5 years of age can remember 92% of things that have seen in form of image as cartoons rather than only read as text such as dialogues of cartoon characters. For example, I do not remember all dialogues of Tom-Jerry Cartoon but I definitely remember how they look and that is because I saw their visual picture. I obtained grasp on this abstraction. It is a reward for me after practicing it manually. I will always recommend to practice manually first then once we get hang on it then can try with other libraries. I am trying many distinct techniques for training my visualization concept clarity because I also prefer viewing and retaining information rather than reading therefore this will also help while presenting in front of company's client and they will have finer perception of it. Hence, I continue to gain knowledge while practicing the same and spread intellectual wings in tech-heaven.

Technical Aspect

Numpy contains a multi-dimensional array and matrix data structures. It works with the numerical data. Numpy is faster because is densely packed in memory due to its homogeneous type. It also frees the memory faster.

Pandas module mainly works with the tabular data. It contains Data Frame and Series. Pandas is 18 to 20 times slower than Numpy. Pandas is seriously a game changer when it comes to cleaning, transforming, manipulating and analyzing data.

Matplotlib is used for EDA. Visualization of graphs helps to understand data in better way than numbers in table format. Matplotlib is mainly deployed for basic plotting. It consists of bars, pies, lines, scatter plots and so on. Inline command display visualization inline within frontends like in Jupyter Notebook, directly below the code cell that produced it.

Installation

Using intel core i5 9th generation with NVIDIA GFORCE GTX1650.

Windows 10 Environment Used.

Already Installed Anaconda Navigator for Python 3.x

The Code is written in Python 3.8.

If you don't have Python installed then please install Anaconda Navigator from its official site.

If you are using a lower version of Python you can upgrade using the pip package, ensuring you have the latest version of pip, *python -m pip install --upgrade pip and press Enter.*

Run/How to Use/Steps

Keep your internet connection on while running or accessing files and throughout too.

Follow this when you want to perform from scratch.

Open Anaconda Prompt, Perform the following steps:

```
cd <PATH>
```

```
pip install numpy
```

```
pip install pandas
```

```
pip install matplotlib
```

You can also create requirement.txt file as, `pip freeze > requirements.txt`
run files.

Follow this when you want to just perform on local machine.

Download ZIP File.

Right-Click on ZIP file in download section and select Extract file option, which will unzip file.

Move unzip folder to desired folder/location be it D drive or desktop etc.

Open Anaconda Prompt, write `cd <PATH>` and press Enter.

eg: `cd C:\Users\Monica\Desktop\Projects\Python Projects 1\12)Visualization\Project_3_Visualization_with_matplotlib`

In Anconda Prompt, `pip install -r requirements.txt` to install all packages.

Open in Jupyter Notebook, `<filename>.ipynb`

That is,

Open in Jupyter Notebook, 1)Visualization of different types of graphs using matplotlib.ipynb

This takes fifa_data.csv and gas_prices files as input dataset.

Please be careful with spellings or numbers while typing filename and easier is just copy filename and then run it to avoid any silly errors.

Note: `cd <PATH>`

[Go to Folder where file is. Select the path from top and right-click and select copy option and paste it next to `cd` one space `<path>` and press enter, then you can access all files of that folder] [`cd` means change directory]

Directory Tree/Structure of Project

Folder: 12)Visualization > Project_3_Visualization_with_matplotlib

1)Visualization of different types of graphs using matplotlib.ipynb

To Do/Future Scope

Can do model building.

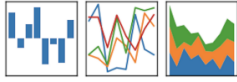
Technologies Used/System Requirements/Tech Stack



NumPy

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



matplotlib

Credits

Pradeep Sir