COLLEGE PLACEMENT
Data Statistic

Data Aanalysis using "PYTHON OPERATIONS"

Guided by::

Mr. Pranav Shriram Sir

Presented by:

Abhijit (821)

Payas(818)

Sejal(806)

INTRODUCTION

Our project revolves around the analysis of college placement data using Python. With Python coding, we performed operations such as finding maximum, minimum, and average salaries, creating informative graphs, and examining branch-wise placement distribution. The aim was to optimize placement strategies and provide valuable insights for students and institutions. Through our project, we demonstrated the power of data analysis and Python in driving informed decision-making for a successful future.

MOTIVATION

During our project, we were driven by curiosity and passion as we analyzed college placement data using Python. We discovered important insights like the highest and lowest salaries and the distribution of placements across different branches. Our motivation stemmed from the desire to make a positive impact, so we shared our findings to inspire others to explore data analysis. This project highlights how data and coding can shape effective strategies for a better future.

Statistics Data Report

- Package Distributions:

Analyze salary packages offered to graduates in 2019-20, 2020-21, and 2022-23.

- Companies Hiring Trends:

Explore trends in companies actively hiring college graduates.

- Year-wise Distribution:

Compare placement rates and trends across the academic years.

Details of Dataset

Let's start by understanding the details of our college placement dataset.

Step 1

We will explore the variables and attributes available, such as student name, branch, Package, and placement status.

Step 2

This dataset will serve as the foundation for our data analysis and statistical exploration.

Pandas

Pandas is a Python library.
Pandas is used to analyze data.

- Pandas provides functions to handle missing data, duplicate values, and outliers. It offers capabilities to clean, transform, and normalize data, making it suitable for further analysis or machine learning tasks.
- Pandas has dedicated functionalities for working with time series data. It provides tools to handle date and time data, resample data at different frequencies, and perform rolling window calculations. These features are valuable for analyzing and modeling time-dependent data.
- Pandas supports various file formats, including CSV, Excel, SQL databases, and more. It enables reading and writing data to and from these formats, making it convenient to import and export data for analysis and sharing.

Example

Using pandas find average salary and then print all the details of that student who got average salary

```
import pandas as pd
data = pd.read_csv('Placement.csv')
avg_package = data['PACKAGE IN LAKH'].mean()
print("Average Package in LAKH:",avg_package)

average = data[(data["PACKAGE IN LAKH"] >= 5) & (data["PACKAGE IN LAKH"] <= 6)]
average
```

Average PACKAGE

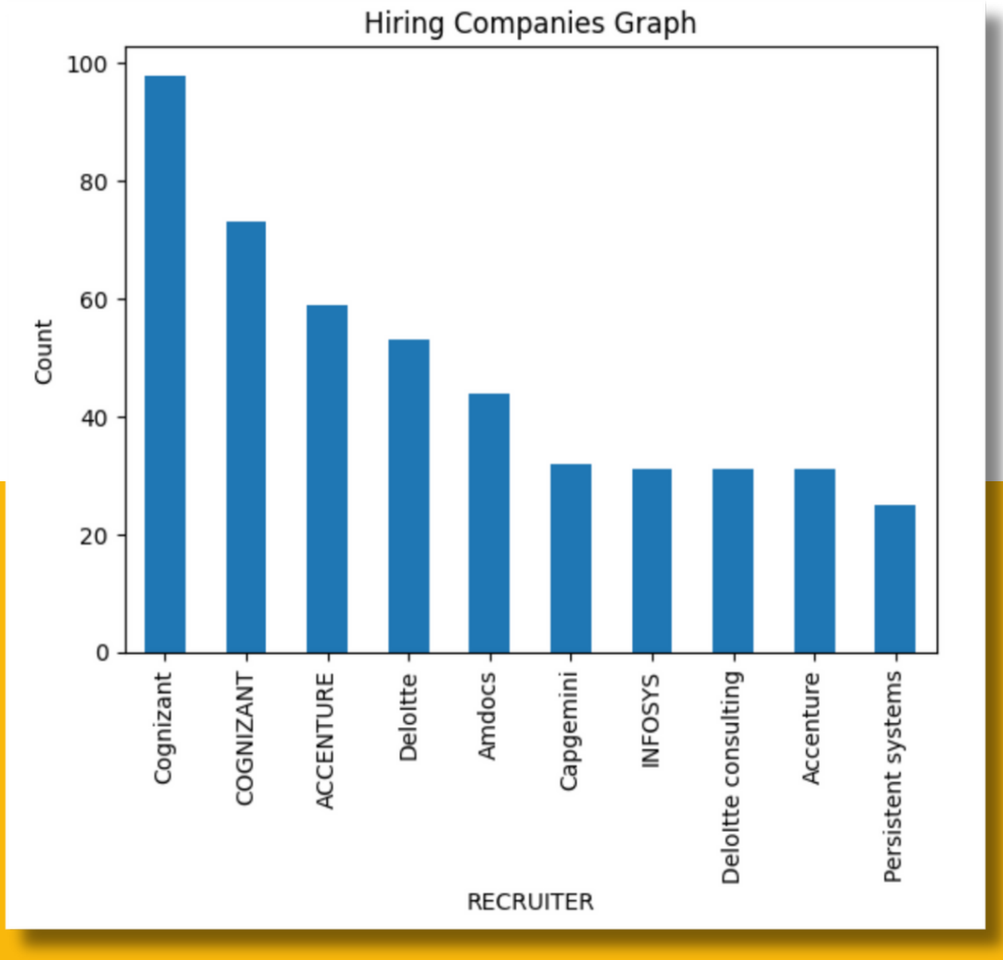
OUTPUT :

	S.NO	NAME OF STUDENT	BRANCH	RECRUItERS	PACKAGE IN LAKH	Year
15	16.0	Aniket Singh	E&TC	Bosch Global Software Technologies Pvt. Ltd	5.0	2022-23
16	17.0	Ayush Sunil Jadhao	ETX	Bosch Global Software Technologies Pvt. Ltd	5.0	2022-23
17	18.0	Dashmesh Udayshankar Singh	Computer	Bosch Global Software Technologies Pvt. Ltd	5.0	2022-23
18	19.0	Prajwal Jagannath Kumbhar	It	Bosch Global Software Technologies Pvt. Ltd	5.0	2022-23
19	20.0	Sahil Jagdish Bhutada	E&TC	Bosch Global Software Technologies Pvt. Ltd	5.0	2022-23
...	...	...	...	...	...	...
1525	NaN	SHARWARI SANDEEP NALAWADE	Mechanical	PRODAIR AIR PRODUCTS INDIA PVT. LTD	5.0	2019-20
1526	NaN	KUNAL SAMADHAN GHATALE	Mechanical	JSW STEEL	6.0	2019-20
1527	NaN	RUTVIJ ASHOK WAKALKAR	Mechanical	JSW STEEL	6.0	2019-20
1528	NaN	AKASH RAJKUMAR PRASAD	Mechanical	JSW STEEL	6.0	2019-20
1579	NaN	YURAJ AHIRE	Mechanical	TCS	5.0	2019-20

224 rows × 6 columns

DATA VISUALIZATION

- Data visualization plays a crucial role in understanding complex patterns and relationships within the dataset.
- We will utilize various visualizations such as bar charts, pie charts to explore different aspects of the placement data.



Here is the example of bar graph that shows which company hire more, less students
Thats the example of data visualisation

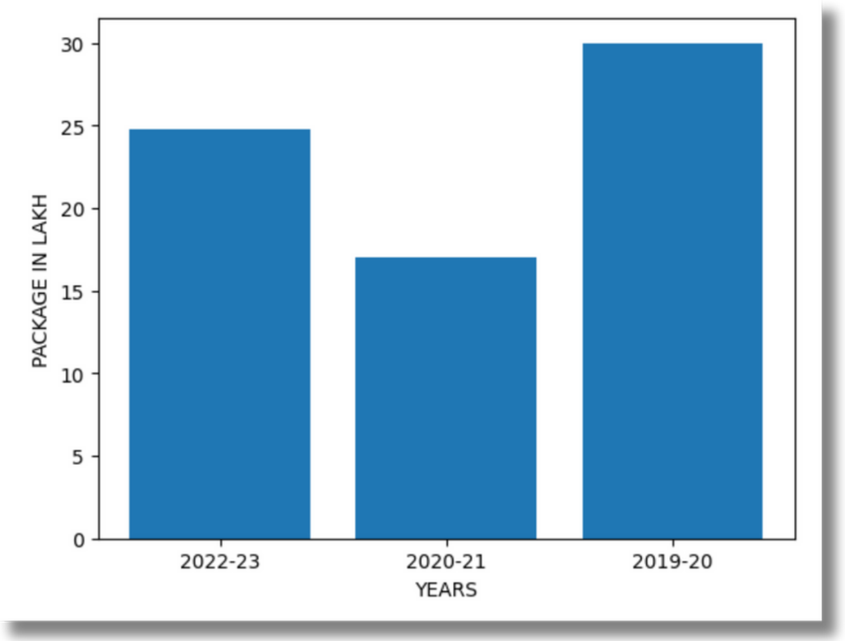
Library used to plot graph:

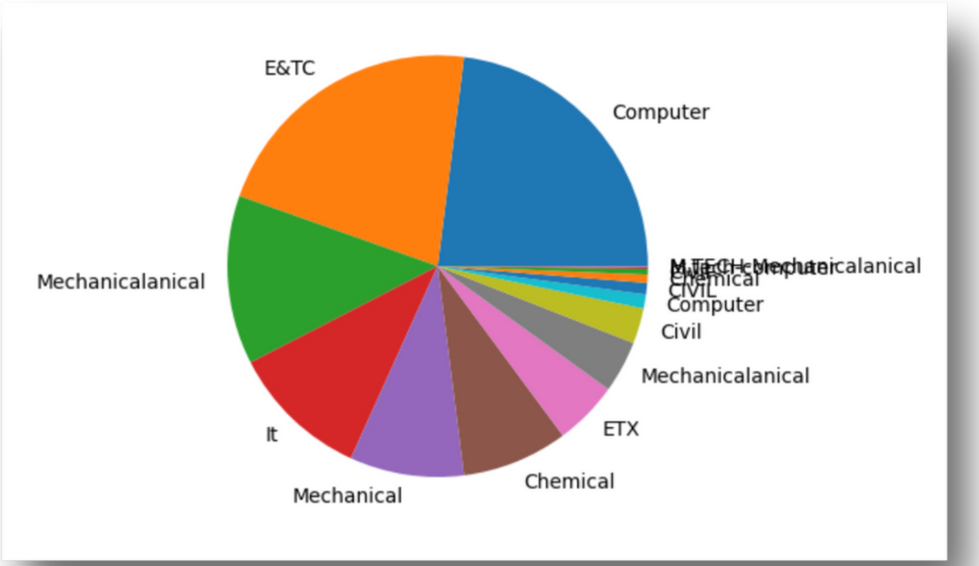
Matplotlib

Matplotlib is a low level graph plotting library in python that serves as a visualization utility.
It is mostly written in python, a few segments are written in C, Objective-C and Javascript
for Platform compatibility...

- Matplotlib provides various functions to create different types of plots, including line plots, scatter plots, bar plots, histograms, and more. It allows you to represent data in a visually appealing manner.
- You can customize almost every aspect of your plot, including colors, line styles, markers, labels, titles, axes, and legends. Matplotlib offers extensive control over the visual appearance of your plots to suit your specific requirements.
- Matplotlib seamlessly integrates with other libraries like NumPy and Pandas, enabling you to directly plot data from these libraries. This integration simplifies the process of visualizing data and streamlines your workflow.

PACKAGE V/S YEAR GRAPH





BRANCH WISE DISTRIBUTION PIE CHART

THANK YOU