REPORT

ON

FOUR WEEKS OF INTERNSHIP-II
Carried out on

"DATA SCIENCE USING PYTHON"

Submitted to

NMAM INSTITUTE OF TECHNOLOGY, NITTE
(An Autonomous Institution under VTU, Belagavi)

In partial fulfillment of the requirements for the award of the

Degree of Bachelor of Engineering
in
Computer Science & Engineering

by

PRASHANTHA NAYAK

4NM22CS410

Under the guidance of

MR. GURUPRASAD

ASSISTANT PROFESSOR GD-III

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NITTE EDUCATION TRUST | NMAM INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the "Internship report-II" submitted by ***Mr. PRASHANTHA NAYAK*** *bearing USN of* ***4NM22CS410*** *V semester B.E. a bonafide student of NMAM Institute of Technology Nitte, has undergone three weeks of internship at* ***BRAINOVISION SOLUTION Pvt Ltd*** *in "****Data Science Using Python****" during SEP 2023 -OCT 2023 fulfilling the partial requirements for the award of degree of Bachelor of Engineering in* ***Computer Science and Engineering*** *at NMAM Institute of Technology, Nitte****.***

Name and Signature of Mentor

Signature of HoD

CERTIFICATE

OF INTERNSHIP COMPLETION

TO

Date: 30-10-2023

Prashantha nayak
N.M.A.M Institute of Technology

This is to certify that **Prashantha nayak** has successfully completed his **ONE MONTH INTERNSHIP** program with **BrainOvision Solutions Pvt. Ltd.** He was working on **DATA SCIENCE** and was actively & diligently involved in the projects and tasks assigned. During the span, we found him punctual and hardworking person.
His feedback and evolution proved that he is a quick learner. Congratulations and Best Wishes.

ROLE	: **DATA SCIENCE INTERN**
INTERN ID	: **BOV23DN242**
START DATE	: **28-09-2023**
END DATE	: **28-10-2023**

Yours Faithfully

Ganesh Nag Boddi
Founder & CEO
Brainovision Solutions India Pvt Ltd

Dr. Buddha Chandrashekar
Chief Coordinating Officer – AICTE
All India Council for Technical Education

ACKNOWLEDGEMENT

The satisfaction that accompanies the completion of any task would be incomplete without the mention of all the people, without whom this endeavor would have been a difficult one to achieve their constant blessings, encouragement, guidance and suggestions have been a constant source of inspiration.

First and foremost, my sincere gratitude to my mentor, **Mr. Guruprasad** Department of Computer Science and Engineering for the constant guidance throughout the course of this internship and for the valuable suggestions.

My sincere thanks to our beloved principal**, Dr. Niranjan N. Chiplunkar** for his support and encouragement

I would also like to thanks **Dr. Jyothi Shetty** -HoD Computer Science and Engineering for giving me an opportunity to prepare for this internship and also sincere thanks for providing all needed facilities.

I would like to thank **Narayana Swamy,** Technical Head and **Anu Sri K** ,Resource team of Brainovision Solutions India Pvt. Ltd, Hyderabad for giving me the opportunity to do this opportunity .I am hugely indebted to our trainer **Mr. Vijay G H** for his immense support and guidance.

I would like to extend my heartfelt gratitude towards all our staff members who have guided and connected me through all possible ways. Last but not the least I want to thank our family and friends who have been a constant support throughout the entire process

Prashantha Nayak
4NM22CS410

TABLE OF CONTETS

SL.NO	*TITLE*	*PAGENO*
1	**ABSTRACT**	**1**
2	**INTRODUCTION TO INDUSTRY**	***2***
3	**DETAILS OF TRAINING**	***3***
4	**DATASCIENCE LIBRARIES**	***4***
5	**PROJECT IMPLEMENTATION**	***5-9***
6	**CONCLUSION**	***10***
7	**REFERENCES**	***11***

ABSTRACT

The primary objective of this internship was to gain practical exposure and apply the theoretical knowledge. During the internship Brainovision team have given the detailed knowledge on the different Python data wrangling and data visualization libraries. I also gained proficiency in data visualization and statistical modeling techniques.

The Internship experience provided me with a comprehensive understanding of the data science workflow and the practical applications like jupyter notebook.I learned to effectively communicate complex data analysis findings to both technical and non-technical solutions such as Data cleaning and preprocessing and Explotary data analysis .This internship enhanced my technical skills and broadened my understanding of data science. I gained hands-on experience with real-world data analysis project.

INTRODUCTION TO THE INDUSTRY

Brainovision Solutions India Private Limited, is an organization which deals with the wing of software development and technical education. This is the place for students and faculty and other companies to find solution for all your requirements Such as internships, academic projects (Mini & Major project), online courses, Workshops, faculty development programs and to hire a perfect skilled candidate.

All the certificate will be issued from our corporate company Brainovision Solutions India Private Limited. If you are the one who dreams to be a technical Pro and wants to get placed in MNCs then this is the place to have to stop and start learning practically in corporate environment. Brainovision Solutions India Private Limited is an organization started in 2014 that serves web solutions, software development, and tech education in the corporate field. This is the place for brands, students, and faculty to find the best solution.

Along with software development and solutions our aim is to fill the gap between academics and corporate. Involve every student in the real time projects and make them corporate ready. Provide an opportunity to engineering students and make them gain exposure to the diverse field. Doing an internship is a great way to gain hands on experience and leave with complete knowledge about the operations in a company It allows students to work with professionals, giving them insights about their future work careers and environment.

Workshops offer structured learning environment that enables individuals to learn engineering technologies and hackathons provide the individuals to face challenge and l earn to work towards solving problem statements of the companies. Job assistance training is a program designed to provide knowledge, support and resources to individuals seeking employment These programs typically offer a variety of technical and employability skills

DEATAILS OF TRAINING UNDERGONE

Introduction

Data science is a field that involves the use of scientific methods, processes, algorithms, and systems to extract meaningful insights and knowledge from structured and unstructured data. It encompasses a wide range of techniques and approaches from statistics

Key component of the Data Science

Data Collection: The first step in any data science project is collecting relevant data. This can involve gathering data from various sources such as databases, APIs, sensors, or even manual data entry.

Data Cleaning and Preprocessing: Raw data often contains errors, missing values, or inconsistencies . Data scientists s pend a significant amount of time cleaning and preprocessing data to ensure its accuracy and reliability.

Exploratory Data Analysis (EDA): EDA involves examining the data visually and statistically to understand its characteristics. Data visualization tools and techniques are often used to uncover patterns, trends, and relationships within the data.

Statistical Analysis: Statistical methods are applied to infer properties of the population from a sample of data. This helps in making data-driven decisions and drawing meaningful conclusions.

Data Visualization: Communicating findings effectively is crucial in data science. Data visualization tools are used to create charts, graphs, and dashboards that make complex data more accessible and understandable.

Data Modeling: Creating mathematical models or algorithms to represent relationships identified in the data, facilitating predictions and insights.

Data Science Libraries

Data scientists use a variety of libraries and frameworks to perform tasks related to data analysis, machine learning, and other aspects of the data science workflow. Here are some commonly used libraries in the Python programming language, which is widely used in the field of data science:

NumPy: A fundamental package for scientific computing with Python. It provides support for large, multi-dimensional arrays and matrices, along with mathematical functions to operate on these arrays. Numpy is a foundational library for numerical and scientific computing in the Python ecosystem, and it serves as the basis for many other libraries in the data science and machine learning domains.

Pandas: A powerful data manipulation and analysis library. It provides data structures like Data Frame for efficient manipulation and analysis of structured data. Th e Data frame is a two-dimensional, tabular data structure with labeled axes (rows and columns). It is similar to a spreadsheet or a SQL table and is the primary data structure e used in Pandas for handling structured data.

Matplotlib: A 2D plotting library for creating static, animated, and interactive visualizations in Python. It is often used for creating charts and graphs to represent data. It is often used in conjunction with other libraries, such as NumPy and Pandas, to create informative and visually appealing graphics.

Seaborn: Built on top of Matplotlib, Seaborn is a statistical data visualization library that provides a high-level interface for drawing attractive and informative statistical graphics

SciPy: A library used for scientific and technical computing. It builds on NumPy and provides additional functionality for optimization, integration, interpolation, eigenvalue problems

PROJECT IMPLEMENATION

Introduction

In the realm of data science, my internship journey unfolded as a quest for practical knowledge and hands-on experience in leveraging data to extract meaningful insights. The rapidly evolving landscape of data analytics prompted me to seek a comprehensive learning environment, and this internship provided the ideal platform to delve into the intricacies of the field

Problem statement

The core challenge addressed in this internship was the analysis of product sales and facing challenges in optimizing its product sales strategy. The company offers a wide range of products on its platform and operates within countries. Despite its global presence, the company is experiencing fluctuations in sales, and there is a need to understand the underlying factors affecting product sales performance. The management believes that leveraging data science techniques can provide valuable insights to enhance product sales, improve customer satisfaction, and ultimately boost revenue.

Methodology

The methodology adopted in this internship encompassed several key stages Initial data exploration involved cleaning and preprocessing diverse datasets related to product sales analysis . Descriptive statistics and visualizations provided an initial understanding of the data's characteristics Leveraged for predictive modeling, aiming to forecast the Tools such as Pandas, NumPy, Matplotlib, and seaborn played pivotal roles in implementing these methodologies. The process was iterative, with continuous refinement based on insights gained at each stage.

Dataset

The dataset for this project includes historical sales data, product details, customer information, pricing data, and information on promotional activities. The data spans multiple years and covers various dimensions, including time, product characteristics, and customer attributes

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt # visualizing data
%matplotlib inline
import seaborn as sns
```

```
# import csv file
df = pd.read_csv('prashantha.csv', encoding= 'unicode_escape')
```

```
df.head(30)
```

	User-ID	NAME	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Proc
0	1002903	Prashant	P00125942	M	23-25	23	0	karnataka	southern	Healthcare	
1	1000732	prakash	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	
2	1001990	Priyanka	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	
3	1001425	pramila	P00237842	F	0-17	16	0	Karnataka	Southern	Construction	
4	1000588	pavan	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	
5	1000588	Joni	P00057942	M	26-35	28	1	Himachal Pradesh	Northern	Food Processing	
6	1001132	Balk	P00018042	F	18-25	25	1	Uttar Pradesh	Central	Lawyer	
7	1002092	Shivangi	P00273442	F	55+	61	0	Maharashtra	Western	IT Sector	

- import pandas as pd: This imports the Pandas library and aliases it as pd for easier reference in the code
- pd.read_csv('prashantha.csv', encoding='unicode_escape'): This function is used to read a CSV file
- The df.head() method in Pandas is used to display the first few rows of a DataFrame. By default, it shows the first 5 rows, different number inside the parentheses to display a different number of rows

```
df.columns
```

```
Index(['User-ID', 'NAME', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount'],
      dtype='object')
```

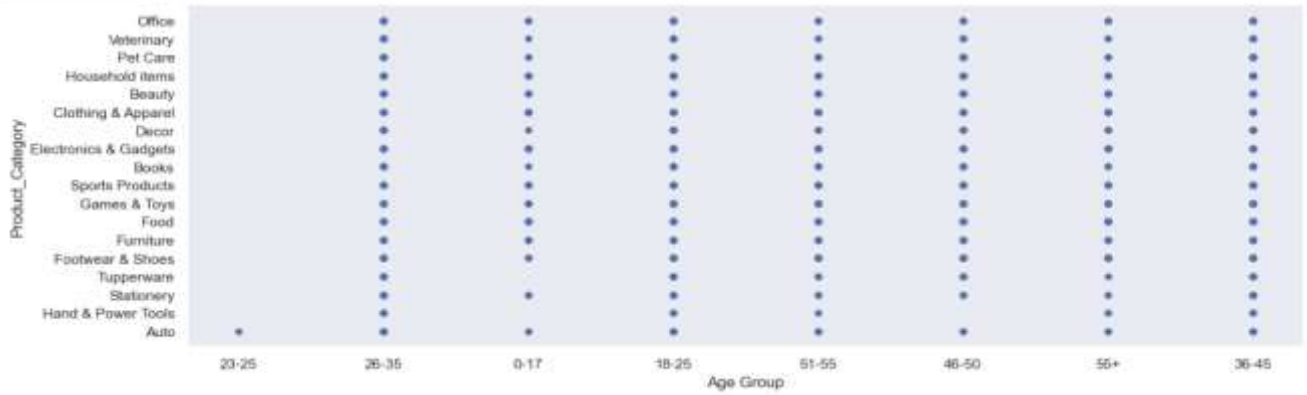
```
df.isnull()
```

	User-ID	NAME	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	O
0	False	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
11246	False	False	False	False	False	False	False	False	False	False	False	
11247	False	False	False	False	False	False	False	False	False	False	False	
11248	False	False	False	False	False	False	False	False	False	False	False	
11249	False	False	False	False	False	False	False	False	False	False	False	
11250	False	False	False	False	False	False	False	False	False	False	False	

- Df.column:specifies various column present in a data frame,It will shows the header Of the columns

- In Pandas, the df.isnull() method is used to check for missing or Nan (Not a Number) values in a Data Frame. It returns a DataFrame of the same shape as the input

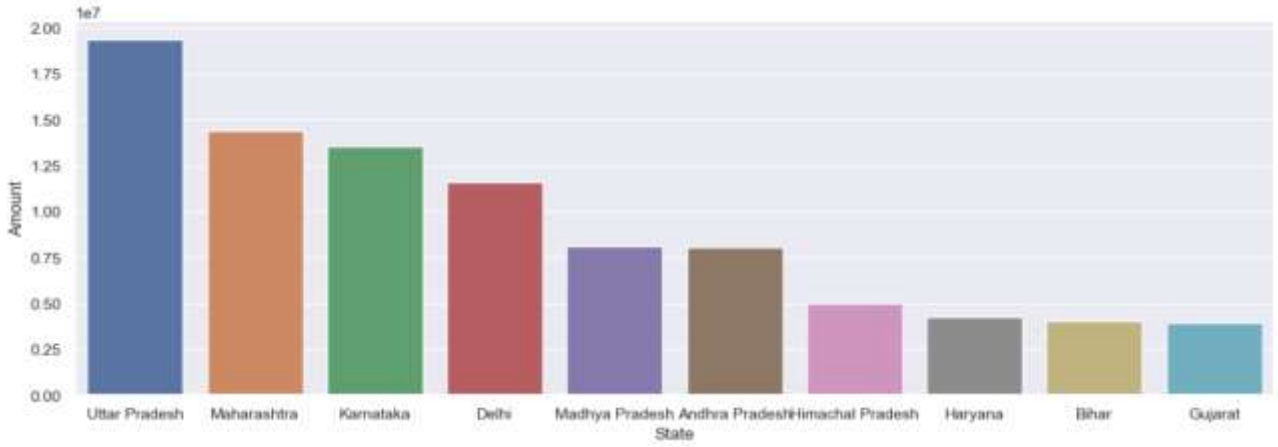
```
df.plot(kind="scatter",
        x="Age Group",
        y="Product_Category")
plt.grid()
```



- This combination of Pandas and Matplotlib allows you to quickly visualize the relationship between "AgeGroup" and "Product_Category" in a scatter plot, with the additional enhancement of grid lines for better readability

```
# total amount/sales from top 10 states

sales_state = df.groupby(['State'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data = sales_state, x = 'State',y= 'Amount')
```

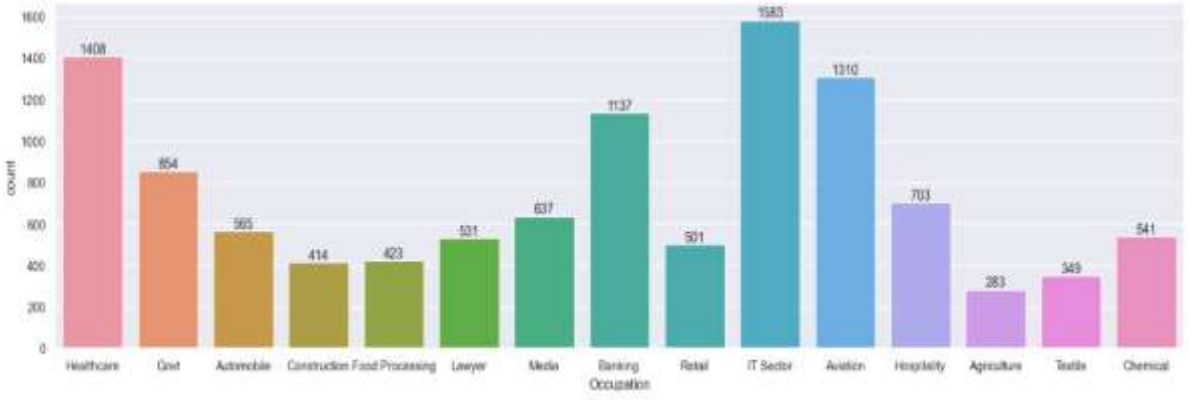
```
<Axes: xlabel='State', ylabel='Amount'>
```



- From above graphs we can see that most of the orders & total sales/amount are from Uttar Pradesh, Maharashtra and Karnataka respectively

Occupation

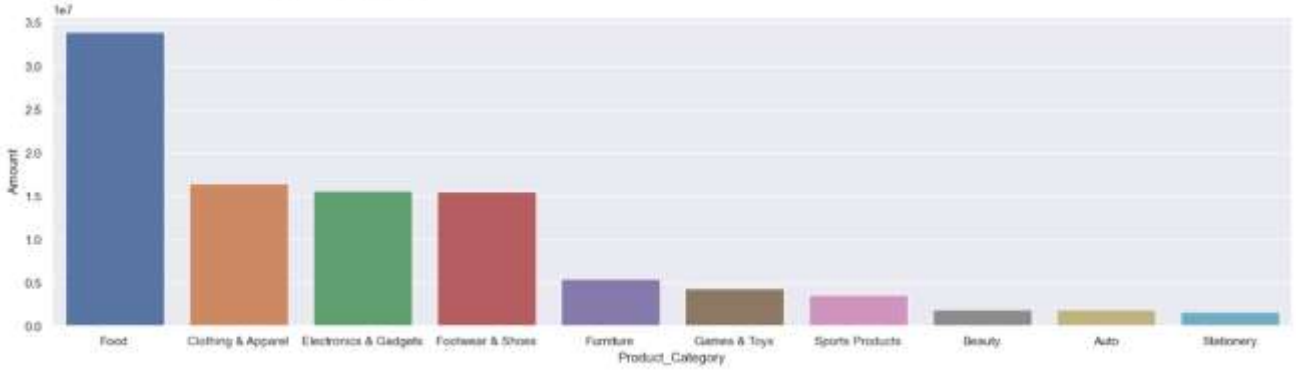
```
#occupation of the various sector
ax = sns.countplot(data = df, x = 'Occupation')
for bars in ax.containers:
    ax.bar_label(bars)
```



- From above graphs we can see that most of the buyers are working in IT, Healthcare and Aviation sector

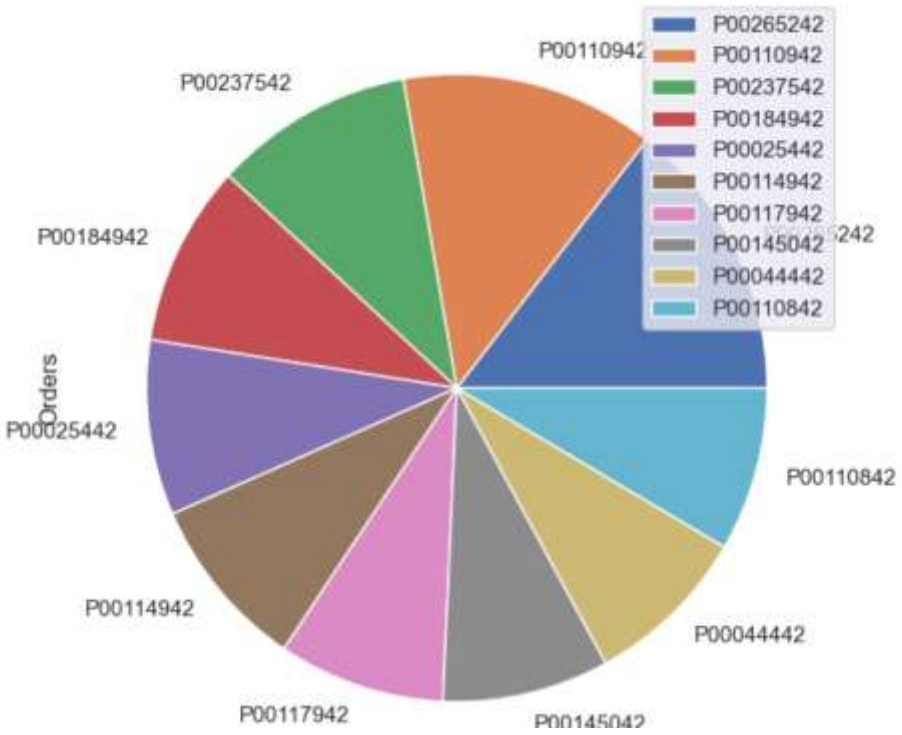
```
sales_state = df.groupby(['Product_Category'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascen

ss.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = sales_state, x = 'Product_Category',y= 'Amount')
```

```
<Axes: xlabel='Product_Category', ylabel='Amount'>
```



- From above graphs we can see that most of the sold products are from Food, Clothing and Electronics category

```
[162]: <matplotlib.legend.Legend at 0x2223073b790>
```



- From the above graph we can see that most of the product sold from the different category

CONCLUSION

This internship strengthened my Python programming skill and expertise in data wrangling and visualization, culminating in data science product analysis project. The project developed using Python and various skillfull use of libraries like Pandas Numpy Matplotlib and Seaborn. Using the different python libraries, implemented the Product Sales Analysis. This project provided valuable insights that can significantly impact sales performance.

REFERENCES

- https://stackoverflow.com/
- https://www.geeksforgeeks.org/
- https://en.m.wikipedia.org/wiki/Data_science
- https://www.w3schools.com/datascience/ds_python.asp
- https://realpython.com/tutorials/data-science/