**CAFE ONE LAST SIP DATA ANALYSIS**

**MAJOR PROJECT REPORT**

***Submitted by***

**ASNEHA SHADMAN**

**2021-310-054**

***In partial fulfillment for the award of the degree of***

**BTECH COMPUTER SCIENCE AND ENGINEERING**

***Under the supervision of***

**DR. SHERIN ZAFAR**

****

**Department of Computer Science and Engineering**

**School of Engineering Science and Technology**

**JAMIA HAMDARD**

**(2024)**

|  |  |  |
| --- | --- | --- |
| **S.No** | **Title** | **Page no.** |
| **1** | **Declaration** | **3** |
| **2** | **Acknowledgement** | **4** |
| **3** | **Title** | **5** |
| **4** | **Objective** | **6** |
| **5** | **Introduction** | **7** |
| **6** | **Problem Statement** | **8** |
| **7** | **Software Requirement Specifications** | **9 -11** |
| **8** | **Project Outline with Snapshots** | **12-17** |
| **9** | **Conclusion** | **18** |
| **10** | **Limitations** | **19** |
| **11** | **Bibliography** | **20** |
| **12** | **End of Project** | **-** |

**DECLARATION**

I, **Asneha Shadman** a student of **‘Btech Cse’** ,Enrolment No: **2021-310-054** , hereby declare that the Project/Dissertation entitled “ **Cafe One Last Sip Data Analysis**” which is being submitted by me to the Department of Computer Science, Jamia Hamdard, New Delhi in partial fulfillment of the requirement for the award of the degree of ‘**Btech Cse**’ is my original work and has not been submitted anywhere else for the award of any Degree, Diploma, Associateship, Fellowship or other similar title or recognition.

**Asneha Shadman**

**Date : 26th Nov ,2024**

**Place : Jamia Hamdard University**

**ACKNOWLEDGEMENT**

I extend my deepest appreciation to my project supervisor , **Dr.Sherin Zafar** ma’am whose unwavering support, insightful feedback, and mentorship have been indispensable throughout this endeavor. Their expertise and encouragement have inspired me to strive for excellence and overcome challenges. I would like to express my gratitude to **Jamia Hamdard University** for providing the resources and environment conducive to research and innovation. I owe a debt of gratitude to my family and friends for their enduring support, understanding, and patience during the demanding yet fulfilling journey of completing this project. Their unwavering belief in my abilities has been a constant source of motivation. This accomplishment would not have been possible without the collective efforts of these individuals and organizations, and for that, I am sincerely thankful.

**Asneha Shadman**

**(Signature and name of the Applicant)**

**TITLE**

**Café One Last Sip DATA ANALYSIS PROJECT**

**OBJECTIVE**

This project focuses on analyzing customer behavior in a coffee shop by simulating data, performing traffic and revenue analysis, and building predictive models. Using Python, it identifies busy times, high-demand items, and revenue drivers through data manipulation and visualization. A Random Forest Classifier predicts customer satisfaction, offering insights to enhance operational efficiency and improve customer experience.

**INTRODUCTION**

This project explores customer behavior within a coffee shop, integrating data analysis, visualization, and predictive modeling to derive actionable insights. Synthetic data representing customer activities—such as time of visit, items purchased, payment method, and satisfaction levels—was created and processed. Using Python libraries, visualizations like heatmaps and bar charts reveal traffic hotspots, peak hours, and revenue-driving items. Additionally, a Random Forest Classifier was implemented to predict customer satisfaction based on transactional and behavioral data. The project demonstrates how data-driven strategies can optimize coffee shop operations, improve customer experience, and enhance decision-making.

**PROBLEM STATEMENT**

Coffee shops often face challenges in understanding customer behavior, managing peak hours, and identifying revenue-driving items, which can lead to inefficiencies in operations and reduced customer satisfaction. There is a need for a comprehensive data-driven solution to analyze customer movement, purchasing patterns, and satisfaction levels. This project aims to address these challenges by simulating customer data, analyzing traffic and sales trends, and building predictive models to optimize operational strategies and enhance customer experience.

**SOFTWARE REQUIREMENTS SPECIFICATION**

I have made this project using Python and its libraries .

Apart from technology if I talk about IDE , I have used VS code for this project.

* For my data analysis project, I utilized Python programming language within the VS Code environment, leveraging its ease of understanding and extensive libraries. Python's global prominence in data science made it the obvious choice, and Jupyter Notebook facilitated an interactive approach to exploring the data. With Python's libraries like pandas for data manipulation and matplotlib for visualization, I conducted a comprehensive analysis, from data cleaning to predictive modeling. Python's versatility and the supportive ecosystem of libraries empowered me to derive actionable insights from the dataset, showcasing its effectiveness in driving data-driven decision-making.

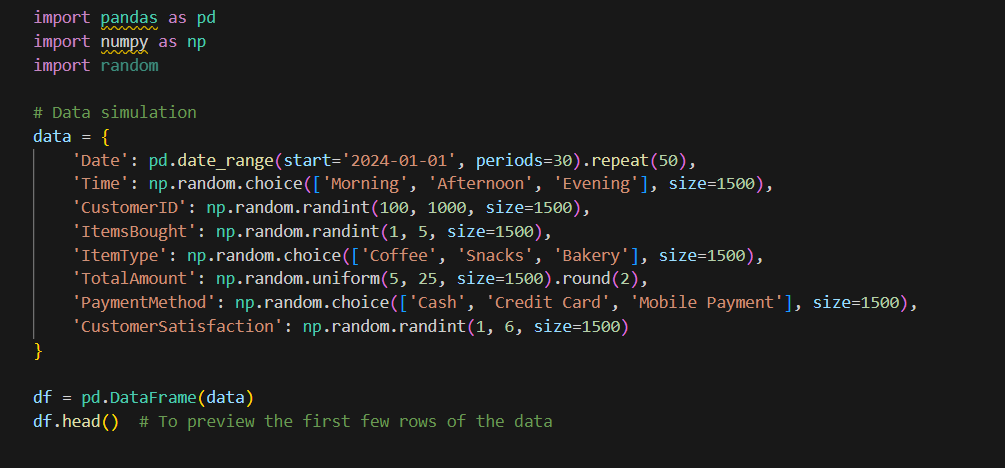


Figure 1Showing Data created in the Project

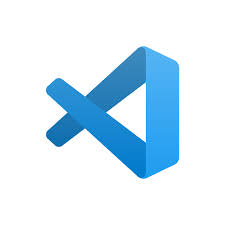
* **Python** is a high-level programming language celebrated for its simplicity and readability. Its extensive standard library and vast ecosystem of third-party packages
* make it versatile for various applications, including web development, data analysis, artificial intelligence, and scientific computing.
* **Libraries of Python :** These libraries play crucial roles in the Python ecosystem, enabling data scientists, analysts, and developers to efficiently manipulate, analyze, and visualize data for various applications. Whether it's exploring datasets, building predictive models, or communicating insights, pandas, NumPy, matplotlib, and seaborn offer powerful tools to streamline the data analysis process.
* seaborn is a statistical data visualization library based on matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics. seaborn simplifies the process of creating complex visualizations, such as categorical plots, distribution plots, and regression plots, with minimal code. It also offers built-in themes and color palettes to enhance the visual appeal of plots.
* matplotlib is a popular plotting library in Python, widely used for creating static, animated, and interactive visualizations. It offers a wide range of plot types, including line plots, scatter plots, histograms, bar charts, and more. matplotlib provides extensive customization options for adjusting plot aesthetics, annotations, and layouts, making it suitable for exploring and presenting data in various formats.
* NumPy is a fundamental library for numerical computing in Python. It provides support for multidimensional arrays, along with a collection of mathematical functions to operate on these arrays efficiently. NumPy is essential for tasks such as linear algebra, Fourier analysis, random number generation, and more. Many other libraries, including pandas, rely on NumPy as a foundational building block.
* pandas is a powerful data manipulation and analysis library in Python. It provides data structures like DataFrame and Series, which enable easy handling of structured data. pandas offers functionalities for tasks such as data cleaning, merging, reshaping, and aggregation. It is widely used in data analysis and preprocessing tasks before applying machine learning algorithms.
* **In my Café One Last Sip Data Analysis Project** : I utilized Python libraries like pandas, NumPy, matplotlib, and seaborn for various tasks. With pandas, I managed the dataset, cleaned missing values, and transformed data types. NumPy facilitated numerical computations and operations, while matplotlib and seaborn were used for visualizing sales trends and relationships between variables. These libraries played essential roles in preprocessing data, analyzing trends, and presenting insights effectively.
* ****I used Jupyter Notebook, a popular interactive computing environment, to conduct my data analysis project in Python. Jupyter Notebook allowed me to write and execute Python code in a web-based interface, combining code, visualizations, and explanatory text in a single document. I created separate cells in the notebook for different parts of my analysis, such as data loading, cleaning, exploration, and visualization. This modular approach made it easy to organize and document my analysis step by step. Additionally, Jupyter Notebook's inline plotting capabilities allowed me to generate and visualize plots directly within the notebook, facilitating iterative .

Figure 2 https://en.wikipedia.org/wiki/Python\_%28programming\_language%29

Figure 3 https://en.wikipedia.org/wiki/Python\_%28programming\_language%29

****

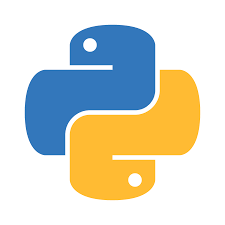
****

Figure 4 https://en.wikipedia.org/wiki/Python\_%28programming\_language%29

**PROJECT OUTLINE WITH SNAPSHOTS:**

**Dataset Overview**

This dataset represents transactional data collected over a month, focusing on customer purchases at a cafe. It includes 1500 entries with the following attributes:

* **Date**: The date of purchase.
* **Time**: The time of the day the transaction occurred (Morning, Afternoon, Evening).
* **CustomerID**: A unique identifier for each customer.
* **ItemsBought**: The number of items purchased in a transaction.
* **ItemType**: The type of items purchased (e.g., Coffee, Snacks, Bakery).
* **TotalAmount**: The total amount spent in a transaction.
* **PaymentMethod**: The method of payment used (Cash, Credit Card, Mobile Payment).
* **CustomerSatisfaction**: A customer satisfaction rating on a scale of 1 to 5.

Below is a snapshot of the dataset showing the first few rows for an overview:

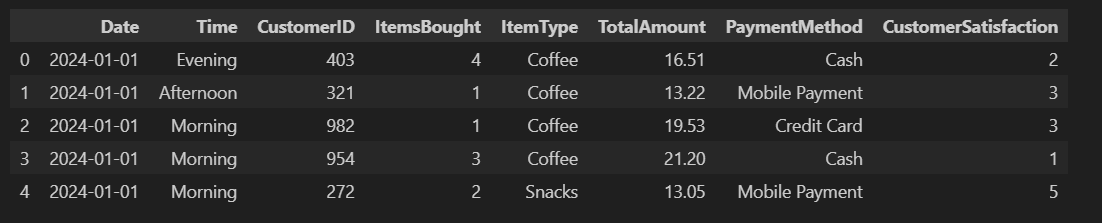


Figure 5 : Output indicating the Data Created

Figure 6 : Plot of distribution for the item types sold

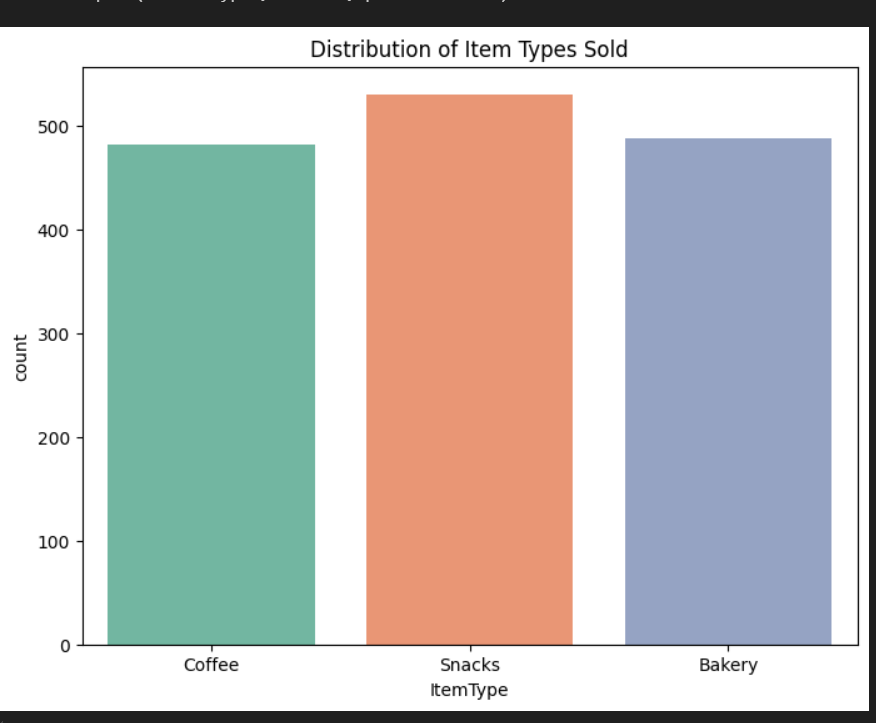
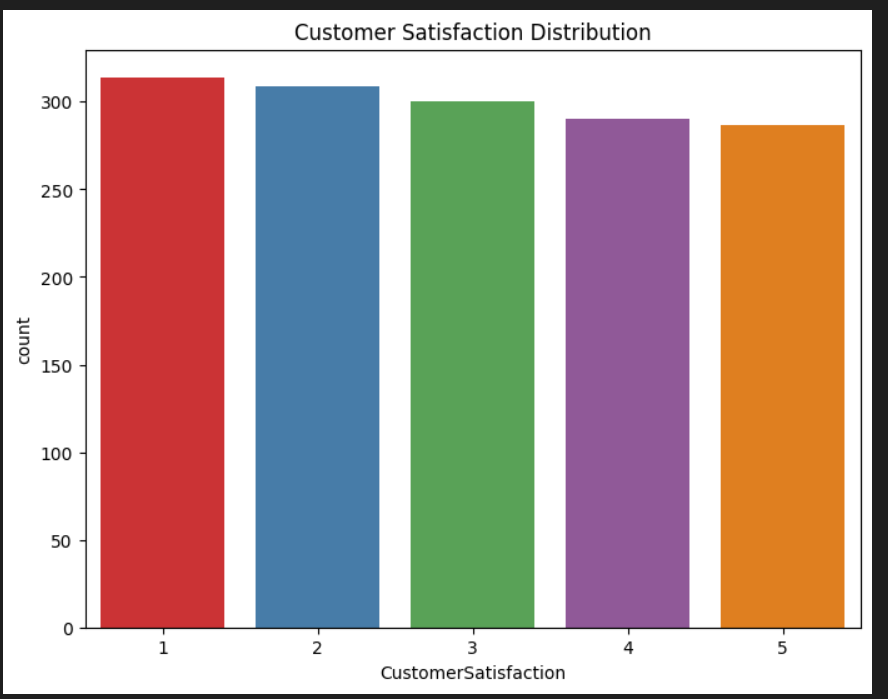


Figure 7 Showing the Distribution of Customer Satisfaction based on the Data



**Total Revenue by Item Type**

The total revenue generated by each item type has been calculated to analyze sales performance. The results show that:

* **Snacks** contributed the highest revenue, indicating their popularity among customers.
* **Bakery items** ranked second in revenue generation.
* **Coffee** had the lowest revenue contribution, likely due to lower item prices or fewer purchases.

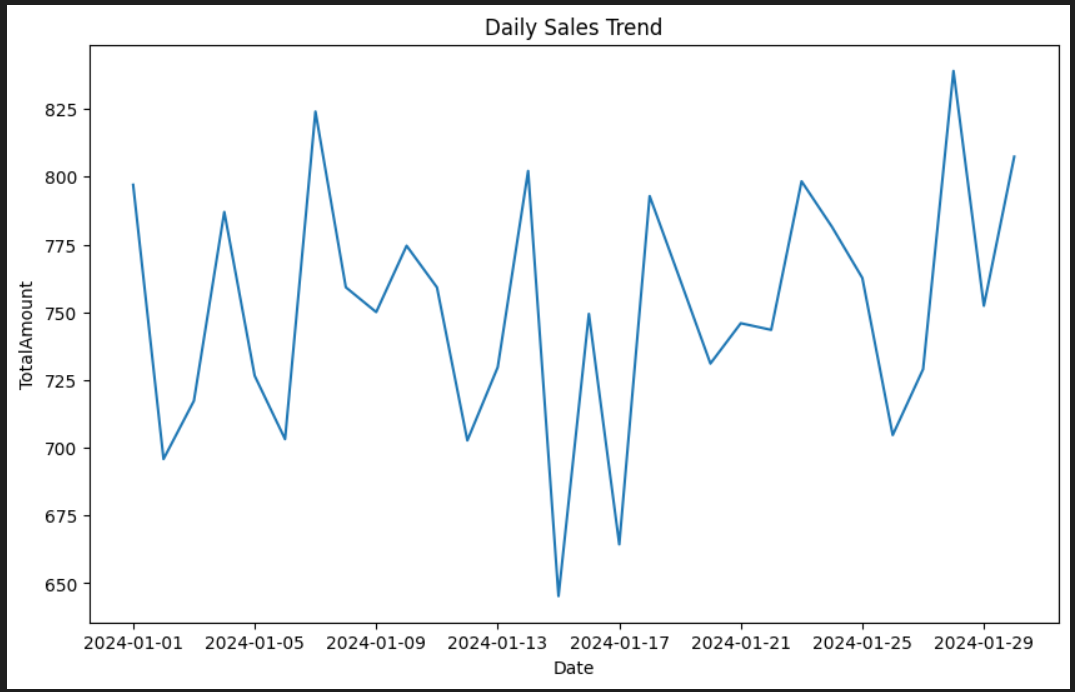


Figure 8 : Shows the Daily Trend Sales Graph

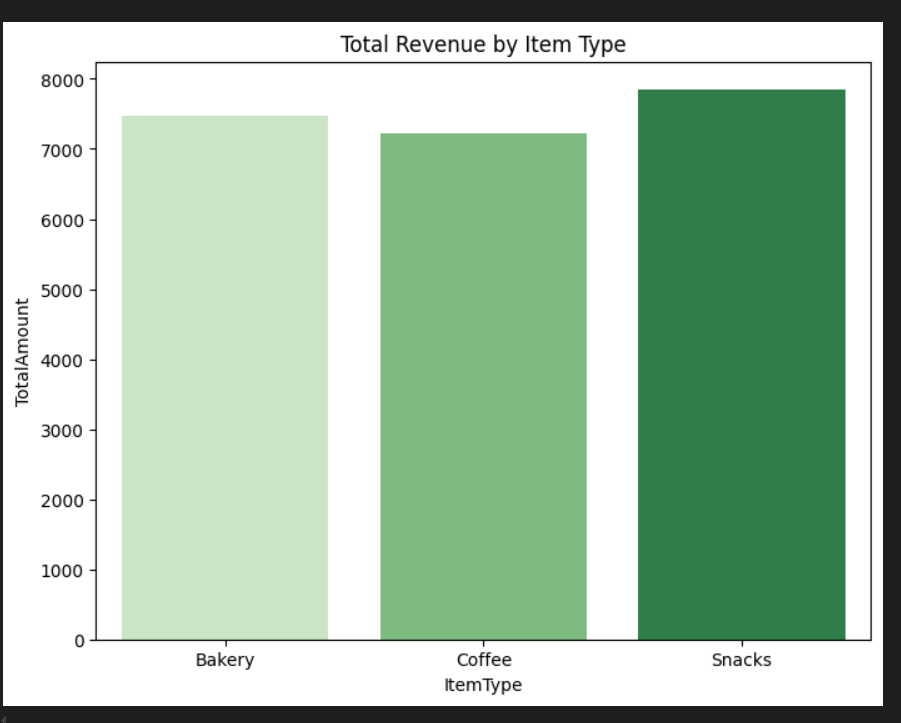


Figure 9 for the output visualization of the trained model.

**Heatmap of Customer Traffic in the Coffee Shop**

A heatmap was generated to analyze customer traffic within different sections of the coffee shop, based on simulated movement data. Key insights include:

* The coffee shop layout is represented as a 5x5 grid, with each section numbered from **Section 1** to **Section 25**.
* The **number of visits** for each section was aggregated and visualized in the heatmap.
* The heatmap highlights the sections with the highest and lowest customer traffic:
  + **Darker shades** represent sections with higher traffic.
  + **Lighter shades** indicate less-visited areas.

This visualization helps identify high-traffic areas, assisting in optimizing layout, product placement, or staffing decisions.

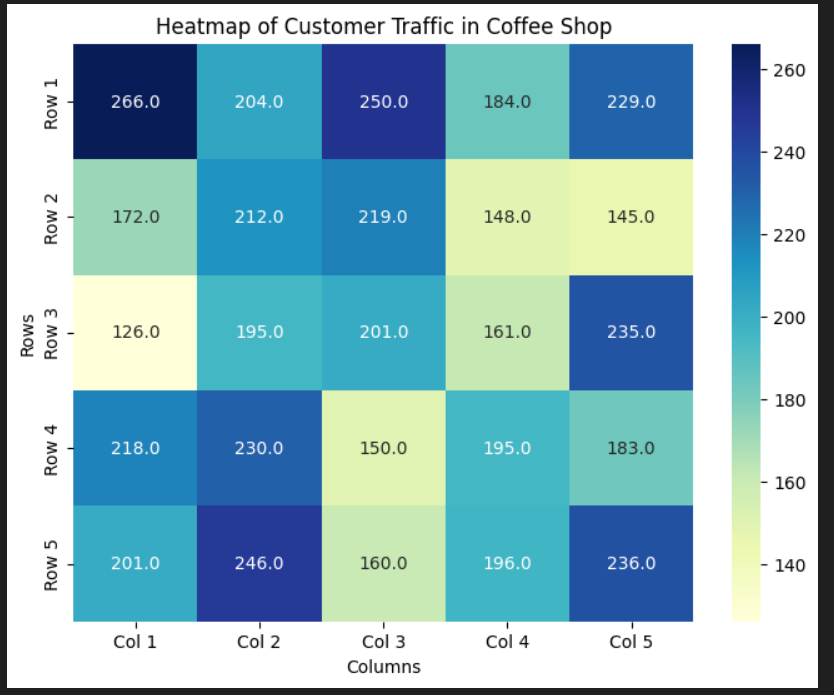


Figure 10 : Shows the concluded result Graph Plotted of Customer Traffic in the Coffee Shop and gives the estimates of highest and lowest traffic

**CONCLUSION**

The analysis of customer behavior and sales patterns in the coffee shop reveals the following key insights:

1. **Customer Traffic Insights**:
   * The section with the **highest number of visits** is **Row 1, Column 1 (266 visits)**, followed by **Row 1, Column 3** and **Row 5, Column 2**.
   * The section with the **lowest traffic** is **Row 3, Column 1 (126 visits)**.
   * Based on these findings, seating arrangements and resources should be optimized, with more seating or offerings relocated to high-traffic areas and underutilized sections re-purposed for better efficiency.
2. **Sales Trends**:
   * **Snacks** were the most purchased items, indicating their popularity, followed by **Bakery items** and then **Coffee**.
   * **Morning and Noon** saw the highest number of purchases, suggesting peak hours of operation, while **Evening** had comparatively lower sales.

These insights provide actionable recommendations to enhance the coffee shop’s layout, inventory management, and operational strategy, ensuring improved customer satisfaction and revenue growth.

**LIMITATIONS**

1. **Simulated Data**:  
   The analysis is based on simulated data, which may not fully represent real-world customer behavior. Actual data collection might reveal different patterns or trends.
2. **Simplified Coffee Shop Layout**:  
   The 5x5 grid layout assumes equal section sizes and uniform traffic distribution, which may not reflect the physical constraints or unique characteristics of a real coffee shop.
3. **Limited Factors Considered**:  
   The analysis focuses primarily on customer traffic, item purchases, and time of day. Other critical factors, such as seasonal variations, promotional impacts, or customer demographics, were not included, potentially limiting the depth of insights.

These limitations highlight areas where real-world data and expanded analysis could provide more accurate and comprehensive results.

**BIBLIOGRAPHY**

**1**.Python Programming language : <https://www.python.org/>

2. <https://seaborn.pydata.org/tutorial/introduction.html>

3. <https://jupyter.org/>

4. Brownlee, J. (2019). *A Gentle Introduction to Machine Learning*. Machine Learning Mastery. Retrieved from <https://machinelearningmastery.com/a-gentle-introduction-to-machine-learning/>

5.How to use Business Analysis in your Project: <https://www.glowtouch.com/business-analysis-process-flow/>

6.Revolutionize your sales strategy with Expert Sales Data Analysis (Blog by Abhishek Punyani , March 21,2024) : <https://convin.ai/blog/sales-data-analysis-tips>