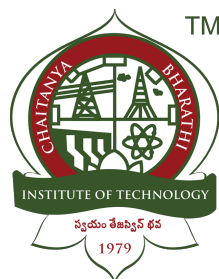


An
Internship Report on
DATA ANALYSIS
submitted in partial fulfillment of the requirements
for
the award of the degree of
Bachelor of Engineering
in
Artificial Intelligence and Data Science

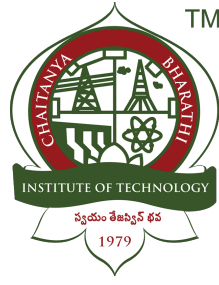
By
Ambekar Tejas (1601-21-771-095)
Under the esteemed guidance of
Mrs. Sheena Mohammad
Assistant Professor
Department of Artificial Intelligence and Data Science



DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE
CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
(AUTONOMOUS)

(Affiliated to Osmania University, Approved by AICTE and Accredited by NAAC)

HYDERABAD – 500075, Telangana, India



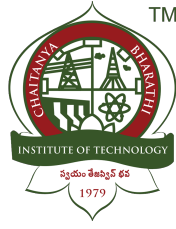
**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE
CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
HYDERABAD – 500075**

DECLARATION CERTIFICATE

I hereby declare that the work described in this report entitled **DATA ANALYSIS** which is being submitted by me in partial fulfillment for the award of **BACHELOR OF ENGINEERING** in the Department of Artificial Intelligence and Data Science, Chaitanya Bharathi Institute of Technology to the Osmania University Hyderabad is original and has not been submitted for any Degree or Diploma of this or any other university.

Project Associate

Ambekar Tejas (1601-21-771-095)



**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE
CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
HYDERABAD – 500075**

BONAFIDE CERTIFICATE

This is to certify that the internship report entitled, **DATA ANALYSIS**, is a bonafide record of the work done by

Ambekar Tejas (1601-21-771-095)

in partial fulfillment of the requirements for the award of the degree of **Bachelor of Engineering in Artificial Intelligence and Data Science** to the **CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY, HYDERABAD** at TwiLearn during the year 2023-24. It is certified that he/she has completed the project satisfactorily.

Signature of Supervisor

Mrs. Sheena Mohammad

Guide

Signature of Head of the Department

Dr.K.Ramana

Associate Professor & Head

INTERNSHIP OFFER LETTER

TwilLearn EduTech Pvt. Ltd.
Ravet, PUNE-412101. India.
Tel: +91 8421796406
www.twilearn.com
CIN: U80900PN2023PTC217646



UNID: CRIN2302047

DATE: 15-06-2023

Dear Ambekar Tejas ,

On behalf of TwilLearn Edutech Pvt. Ltd. , I am pleased to extend to you this offer of temporary employment as a Data Science Intern for a 3-months duration. If you accept this offer, you will begin your internship with the Company. As an intern, you will receive "temporary employment" status. As a temporary employee, you will not receive any of the employee benefits that regular company employees receive.

This letter constitutes the complete understanding between you and the Company regarding your internship and supersedes all prior discussions or agreements. This letter may only be modified by a written agreement signed by both of us. I hope that your internship with the Company will be successful and rewarding.

Sincerely,

Archana S.

Archana Swami
Director/Co-founder
TwilLearn EduTech Pvt. Ltd.

 **Phone**
+91 8121796406

 **Email**
hello@twilearn.com

 **Address**
Ravet, PUNE-412101

INTERNSHIP CERTIFICATE

CIN:U80900PN2023PTC217646

UNID: CRIN2302047



CERTIFICATE

OF INTERNSHIP

THIS CERTIFICATE IS PRESENTED TO

Ambekar Tejas

For Successfully completing 3-months Internship as a Data Science Intern at TwilLearn Edutech Pvt. Ltd. During this Internship we found him/her consistent & hardworking. We wish him/her all the best for future endeavors.

Archana S.

ARCHANA SWAMI
DIRECTOR



Onkar Mathapati

ONKAR MATHAPATI
FOUNDER/CEO

ISSUED ON: 13/09/2023

ACKNOWLEDGEMENTS

The satisfaction that accompanies the successful completion of the task would be put incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crown all the efforts with success.

The organization TwiLearn and our mentor **SaiPrasad Kagne** is very helpful and continuously encourage me to complete the task assigned by the company. Organization support is also very good. They provided all the basic requirements required to complete the task.

We express our heartfelt thanks to **Mrs. Sheena Mohammad**, Assistant Professor, for his / her suggestions in the selection and carrying out an in-depth study of the topic. His / Her valuable guidance and encouragement really helped us to shape this report to perfection.

We express our heartfelt thanks to **Mrs.S.Shoba Rani**, Assistant Professor & Internship Coordinator and **Dr.P.Srilatha**, Associate Professor & Project Coordinator, for her suggestions invaluable inputs and assessment really helped us to shape this report to perfection.

We wish to express our deep sense of gratitude to **Dr. Kadiyala Ramana**, Associate Professor & Head, for his intense support and encouragement, which helped us to mold our project into a successful one.

We also owe our special thanks to our honorable Principal **Dr. C. Venkata Narasimhulu**, for providing all the infrastructural facilities and congenial atmosphere to complete the project successfully.

We also thank all the staff members of the Artificial Intelligence and Data Science department for their valuable support and generous advice.

Finally, thanks to all my friends and family members for their continuous support and enthusiastic help.

ABSTRACT

Data Science, a multidisciplinary field, transforms data into actionable insights crucial for decision-making and innovation. Its importance lies in revealing hidden patterns and trends, fostering a data-driven approach. Data analysis, a fundamental tool, enables organizations to glean valuable information, optimizing processes and enhancing overall success in today's data-rich landscape.

Twilearn EduTech Pvt. Ltd. stands out with its practical approach, offering one-on-one mentorship and project-oriented training and internship programs in Data Science, Web Development, and HR Management. With a track record of 400+ completed projects, including 500+ live sessions and over 50 hours of pre-recorded content, the platform's internship programs not only involve real-life projects but also offer resume building and interview preparation sessions. Emphasis is placed on real-life projects during the internship. Twilearn's commitment to practical skills, mentorship, and support makes it a valuable hub for talent with intelligence.

During my internship at Twilearn, I immersed myself in a comprehensive Data Science program, gaining theoretical knowledge and hands-on experience in Python programming and Tableau Software for data analysis and interpretation.

One significant project focused on **Heart Disease Diagnostic Analysis**. Using Python libraries like matplotlib and seaborn, I applied statistical and machine learning techniques to healthcare data. Exploratory Data Analysis (EDA) revealed crucial patterns, emphasizing the importance of precision in healthcare data analysis.

In the **Swiggy Data Analysis** project, I explored customer behavior and ordering trends using Python libraries for data manipulation and visualization. This project underscored the practical applications of Data Science in optimizing business strategies and operations.

A key aspect of the internship was creating a **Tableau** dashboard, integrating findings for both projects. The dashboard served as a powerful tool for data communication, showcasing my ability to translate complex analyses into impactful visual narratives.

TABLE OF CONTENTS

| Title | Page No. |
|--|----------|
| ACKNOWLEDGEMENTS | i |
| ABSTRACT | ii |
| TABLE OF CONTENTS | iii |
| LIST OF TABLES | vii |
| LIST OF FIGURES | ix |
| Weekly Report - Week 1 | x |
| Weekly Summary - Week 1 | xi |
| Weekly Report - Week 2 | xii |
| Weekly Summary - Week 2 | xiii |
| Weekly Report - Week 3 | xiv |
| Weekly Summary - Week 3 | xv |
| Weekly Report - Week 4 | xvi |
| Weekly Summary - Week 4 | xvii |
| Weekly Report - Week 5 | xviii |
| Weekly Summary - Week 5 | xix |
| Weekly Report - Week 6 | xx |
| Weekly Summary - Week 6 | xxi |
| Weekly Report - Week 7 | xxii |
| Weekly Summary - Week 7 | xxiii |
| CHAPTER 1 INTRODUCTION | 1 |
| 1.1 INTRODUCTION | 1 |
| 1.2 SCOPE | 1 |

| | | |
|------------------|---|----------|
| 1.3 | PURPOSE / OBJECTIVE OF THE PROJECTS | 1 |
| 1.4 | PROBLEM STATEMENT | 2 |
| 1.4.1 | Heart Disease Diagnostic Analysis: | 2 |
| 1.4.2 | Swiggy Data Analysis: | 2 |
| 1.5 | LIMITATIONS | 2 |
| 1.5.1 | Heart Disease Diagnostic Analysis: | 2 |
| 1.5.2 | Swiggy Data Analysis: | 3 |
| CHAPTER 2 | LITERATURE SURVEY | 4 |
| 2.1 | Paper-1: Predictions in heart disease using techniques of data mining by Monika Gandhi,Computer Science and Engineering Department ASET, Amity University, Noida, India; Shailendra Narayan Singh, Amity University, Noida, Uttar Pradesh, India[1] | 4 |
| 2.2 | Paper-2: Using data mining techniques in heart disease diagnosis and treatment by Mai Shouman; Tim Turner; Rob Stocker, School of Engineering and Information Technology, Australian Defence Force Academy, University of New South Wales, Canberra, ACT, Australia[2] | 4 |
| 2.3 | Paper-3: An Analysis of Online Food Ordering Applications in In- dia: Zomato and Swiggy, Anupriya Saxena, Amity University, ABS, Lucknow, Uttar Pradesh, India[3] | 5 |
| 2.4 | Paper-4: Food Delivery Services and Customer Preference: A Com- parative Analysis by Natarajan Chandrasekhar,Saloni Gupta & Nam- rata Nanda[4] | 5 |
| CHAPTER 3 | ANALYSIS | 7 |
| 3.1 | INTRODUCTION | 7 |
| 3.2 | REQUIREMENTS | 7 |
| 3.2.1 | FUNCTIONAL REQUIREMENTS | 7 |
| 3.2.2 | NON-FUNCTIONAL REQUIREMENTS | 7 |
| 3.2.3 | SOFTWARE REQUIREMENTS | 7 |
| 3.2.4 | HARDWARE REQUIREMENTS | 8 |

| | | |
|------------------|---|-----------|
| 3.3 | FLOWCHART | 8 |
| CHAPTER 4 | SYSTEM DESIGN OR METHODOLOGY | 9 |
| 4.1 | INTRODUCTION | 9 |
| 4.2 | DATA FLOW DIAGRAM | 9 |
| 4.3 | MODULE DESIGN AND ORGANSIATION | 10 |
| 4.3.1 | Heart Disease Diagnostic Analysis | 10 |
| 4.3.2 | Swiggy Data Analysis | 11 |
| CHAPTER 5 | IMPLEMENTATION | 13 |
| 5.1 | HEART DISEASE DIAGNOSTIC DATA ANALYSIS | 13 |
| 5.1.1 | Data Understanding and Preparation: | 13 |
| 5.1.2 | Exploratory Data Analysis (EDA): | 13 |
| 5.1.3 | Feature Engineering: | 13 |
| 5.1.4 | Visualisation | 13 |
| 5.1.5 | Tableau Dashboard Creation: | 13 |
| 5.1.6 | Interpretation and Insights: | 14 |
| 5.2 | SWIGGY- BANGALORE OUTLET DATA ANALYSIS | 14 |
| 5.2.1 | Data Collection and Cleaning: | 14 |
| 5.2.2 | Exploratory Data Analysis (EDA): | 14 |
| 5.2.3 | Order Analysis: | 14 |
| 5.2.4 | Customer Segmentation: | 14 |
| 5.2.5 | Visualisation | 14 |
| 5.2.6 | Tableau Dashboard Creation: | 15 |
| 5.2.7 | Insights and Presentation: | 15 |
| CHAPTER 6 | RESULTS | 16 |
| 6.1 | HEART DISEASE DIAGNOSTIC ANALYSIS | 16 |
| 6.2 | SWIGGY - BANGALORE OUTLET DATA ANALYSIS | 17 |
| CHAPTER 7 | FUTURE SCOPE | 18 |
| 7.1 | HEART DISEASE DIAGNOSTIC ANALYSIS | 18 |
| 7.2 | SWIGGY DATA ANALYSIS | 18 |

| | |
|-----------------------------|-----------|
| CHAPTER 8 CONCLUSION | 20 |
| BIBILOGRAPHY | 21 |

LIST OF TABLES

| | | |
|----|---|------|
| 1 | Tasks for Week 1 | x |
| 2 | Supervisor Remarks for Week 1 | x |
| 3 | Work assigned for Week 1 | xi |
| 4 | Learning Outcomes for Week 1 | xi |
| 5 | Remarks for Week 1 | xi |
| 6 | Tasks for Week 2 | xii |
| 7 | Supervisor's Remarks for Week 2 | xii |
| 8 | Work assigned for Week 2 | xiii |
| 9 | Learning Outcomes for Week 2 | xiii |
| 10 | Remarks for Week 2 | xiii |
| 11 | Tasks for Week 3 | xiv |
| 12 | Supervisor's Remarks for Week 3 | xiv |
| 13 | Work assigned for Week 3 | xv |
| 14 | Learning Outcomes for Week 3 | xv |
| 15 | Remarks for Week 3 | xv |
| 16 | Tasks for Week 4 | xvi |
| 17 | Supervisor's Remarks for Week 4 | xvi |
| 18 | Work assigned for Week 4 | xvii |
| 19 | Learning Outcomes for Week 4 | xvii |
| 20 | Remarks for Week 4 | xvii |

| | | |
|----|---|-------|
| 21 | Tasks for Week 5 | xviii |
| 22 | Supervisor's Remarks for Week 5 | xviii |
| 23 | Work assigned for Week 5 | xix |
| 24 | Learning Outcomes for Week 5 | xix |
| 25 | Remarks for Week 5 | xix |
| 26 | Tasks for Week 6 | xx |
| 27 | Supervisor's Remarks for Week 6 | xx |
| 28 | Work assigned for Week 6 | xxi |
| 29 | Learning Outcomes for Week 6 | xxi |
| 30 | Remarks for Week 6 | xxi |
| 31 | Tasks for Week 7 | xxii |
| 32 | Supervisor's Remarks for Week 7 | xxii |
| 33 | Work assigned for Week 7 | xxiii |
| 34 | Learning Outcomes for Week 7 | xxiii |
| 35 | Remarks for Week 7 | xxiii |

LIST OF FIGURES

| | | |
|-----|--|----|
| 3.1 | Flowchart of the Data Analysis and Visualization Process | 8 |
| 4.1 | DFD | 9 |
| 4.2 | Heart Disease Diagnosis Dataset | 10 |
| 4.3 | Swiggy-Bangalore Outlet Dataset | 11 |
| 6.1 | Visualizations created for Heart Disease Diagnostic Analysis | 16 |
| 6.2 | Tableau Dashboard created for the Heart Disease Diagnostic Analysis | 16 |
| 6.3 | Visualizations created for Swiggy Data Analysis | 17 |
| 6.4 | Tableau Dashboard created for the Swiggy Data Analysis | 17 |

Weekly Report - Week 1

| S.No. | Date | Description of Work | Sign of Supervisor |
|-------|------------|--|--------------------|
| 1 | 18-06-2023 | Induction Session | |
| 2 | 19-06-2023 | Python Basics – Part 1 : Introduction, Programming Importance, Variables & Operators. | |
| 3 | 20-06-2023 | Python Basics – Part 2 : List Methods, String Methods, Conditional Statements. | |
| 4 | 21-06-2023 | Python Basics – Part 3 : Loops, Functions, Other Must Know Functions. | |
| 5 | 22-06-2023 | Python Data Structures : Tuples, Sets, Dictionaries and their Methods, Comprehensions | |
| 6 | 23-06-2023 | Python Basics Assignment : Questions based on Python basics and Data Structures concepts. | |

Table 1: Tasks for Week 1

| | |
|--------------------------------|-----------------|
| Name of the Supervisor: | Saiprasad Kagne |
| Supervisor's Sign: | |
| Date: | 24-06-2023 |
| Remarks by Supervisor: | |

Table 2: Supervisor Remarks for Week 1

Weekly Summary - Week 1

Duration: From: 18-06-2023 To: 24-06-2023

Work Assigned:

| |
|--------------------------|
| Python Basics Assignment |
|--------------------------|

Table 3: Work assigned for Week 1

Learning Outcomes:

| |
|---|
| Foundational understanding of Python basics. |
| Proficiency in Python data structures (tuples, sets, dictionaries). |
| Enhanced problem-solving skills through practical assignments. |
| Application of Python programming in real-world scenarios. |
| Appreciation for the broader importance of programming. |

Table 4: Learning Outcomes for Week 1

Remarks:

| |
|--|
| Conducted induction session and initiated Python basics learning. |
| Explored Python data structures and reinforced concepts through assignments. |
| Applied Python programming skills to solve practical challenges. |
| Gained insights into the broader significance of programming in diverse domains. |

Table 5: Remarks for Week 1

Name of the Supervisor: SaiPrasad Kagne

Sign of the Supervisor:

Weekly Report - Week 2

| S.No. | Date | Description of Work | Sign of Supervisor |
|-------|------------|--|--------------------|
| 1 | 26-06-2023 | Numpy – Part 1 : Array Dimensions, List vs Numpy, Reshaping, Indexing, Operations. | |
| 2 | 27-06-2023 | Numpy – Part 2 : View vs Copy, Hstack vs Vstack, Concatenation, Insert Append Delete Functions. | |
| 3 | 28-06-2023 | Pandas – Part 1 : Pandas Series, DataFrame, Concatenation, Top Commands, Indexing DataFrame. | |
| 4 | 29-06-2023 | Pandas – Part 2 : Upload and Read Data, Groupby, Data Range, Reading CSV files. | |
| 5 | 30-06-2023 | Python Library Assignment : Questions based on Numpy and Pandas libraries. | |

Table 6: Tasks for Week 2

| | |
|--------------------------------|-----------------|
| Name of the Supervisor: | Saiprasad Kagne |
| Supervisor's Sign: | |
| Date: | 01-07-2023 |
| Remarks by Supervisor: | |

Table 7: Supervisor's Remarks for Week 2

Weekly Summary - Week 2

Duration: From: 26-06-2023 To: 01-07-2023

Work Assigned:

| |
|-----------------------------------|
| Python Data Structures Assignment |
|-----------------------------------|

Table 8: Work assigned for Week 2

Learning Outcomes:

| |
|--|
| Proficiency in Numpy for array manipulation and indexing. |
| Mastery of advanced Numpy and Pandas functionalities. |
| Efficient use of Pandas for data manipulation and analysis. |
| Practical application of Numpy and Pandas in a library assignment. |

Table 9: Learning Outcomes for Week 2

Remarks:

| |
|--|
| Advanced Numpy skills acquired for array manipulation. |
| Efficient data manipulation using Pandas. |
| Applied Numpy and Pandas in real-world tasks. |
| Strengthened data handling skills for future analysis. |

Table 10: Remarks for Week 2

Name of the Supervisor: SaiPrasad Kagne

Sign of the Supervisor:

Weekly Report - Week 3

| S.No. | Date | Description of Work | Sign of Supervisor |
|-------|------------|---|--------------------|
| 1 | 03-07-2023 | Data Cleaning Module : Introduction, Nan Cases, Missing Value Treatment, Ffill, Bfill, Imputation Techniques, DataFrame Iterations, Pandas Function. | |
| 2 | 04-07-2023 | Exploratory Data Analysis – Part 1 : View vs Copy, Hstack vs Vstack, Concatenation, Insert Append Delete Functions. | |
| 3 | 05-07-2023 | Exploratory Data Analysis – Part 2: IQR Outlier & Z Score, Data Visualization, Complete Matplotlib and Seaborn libraries. | |
| 4 | 06-07-2023 | Python EDA Assignment : Questions based on Data Cleaning & Exploratory Data Analysis concepts. | |

Table 11: Tasks for Week 3

| | |
|--------------------------------|-----------------|
| Name of the Supervisor: | Saiprasad Kagne |
| Supervisor's Sign: | |
| Date: | 08-07-2023 |
| Remarks by Supervisor: | |

Table 12: Supervisor's Remarks for Week 3

Weekly Summary - Week 3

Duration: From: 03-07-2023 To: 08-07-2023

Work Assigned:

| |
|-----------------------|
| Python EDA Assignment |
|-----------------------|

Table 13: Work assigned for Week 3

Learning Outcomes:

| |
|---|
| Proficiency in data cleaning techniques, including handling NaN cases, missing value treatment, and imputation using Pandas functions. |
| Advanced skills in Exploratory Data Analysis (EDA), encompassing various Pandas functionalities and data visualization techniques using Matplotlib and Seaborn. |
| In-depth understanding of outlier detection methods like IQR and Z Score. |
| Practical application of learned concepts through the Python EDA Assignment. |

Table 14: Learning Outcomes for Week 3

Remarks:

| |
|---|
| Mastered data cleaning methods, enhancing data quality. |
| Explored advanced EDA techniques and data visualization libraries. |
| Gained proficiency in outlier detection using IQR and Z Score. |
| Applied data cleaning and EDA skills in a comprehensive assignment. |

Table 15: Remarks for Week 3

Name of the Supervisor: SaiPrasad Kagne

Sign of the Supervisor:

Weekly Report - Week 4

| S.No. | Date | Description of Work | Sign of Supervisor |
|-------|------------|---|--------------------|
| 1 | 10-07-2023 | Statistics – Part 1 : Introduction, Mean, Median, Mode, Population & Sample Mean. | |
| 2 | 11-07-2023 | Statistics – Part 2 : Population & Sample Variance, Standard Deviation. | |
| 3 | 12-07-2023 | Stats Behind Plots – Part 1 : Univariate, Bivariate, Multivariate, Histogram, Box-plot, Barplot, etc. | |
| 4 | 13-07-2023 | Stats Behind Plots – Part 2 : Scatterplot, Line Plot, Pie Plot, Treemap, Advanced Line, Bar & Pie Charts, etc. | |

Table 16: Tasks for Week 4

| | |
|--------------------------------|-----------------|
| Name of the Supervisor: | Saiprasad Kagne |
| Supervisor's Sign: | |
| Date: | 15-07-2023 |
| Remarks by Supervisor: | |

Table 17: Supervisor's Remarks for Week 4

Weekly Summary - Week 4

Duration: From: 10-07-2023 To: 15-07-2023

Work Assigned:

| |
|---|
| Apply statistical concepts learned to analyze a dataset and draw basic plots. |
|---|

Table 18: Work assigned for Week 4

Learning Outcomes:

| |
|---|
| Comprehensive understanding of basic statistics, including mean, median, mode, population mean, and sample mean. |
| Proficiency in statistical measures of variance and standard deviation for both population and sample data. |
| Exploration of statistical foundations behind univariate, bivariate, and multivariate plots, including Histogram, Boxplot, and Barplot. |
| Mastery of advanced plotting techniques, such as Scatterplot, Line Plot, Pie Plot, Treemap, and advanced variations of Line, Bar, and Pie Charts. |

Table 19: Learning Outcomes for Week 4

Remarks:

| |
|--|
| Acquired foundational statistical knowledge for data analysis. |
| Mastered statistical measures of variance and standard deviation. |
| Explored statistical insights behind diverse univariate and bivariate plots. |
| Gained advanced plotting skills, enhancing data visualization capabilities. |

Table 20: Remarks for Week 4

Name of the Supervisor: SaiPrasad Kagne

Sign of the Supervisor:

Weekly Report - Week 5

| S.No. | Date | Description of Work | Sign of Supervisor |
|-------|----------------------------|---|--------------------|
| 1 | 17-07-2023 & 18-07-2023 | Statistics Case Study 1 : Concepts of Measure of Central Tendency. | |
| 2 | 19-07-2023 & 20-07-2023 | Statistics Case Study 2 : Concepts of Descriptive Statistics working with Employees dataset. | |
| 3 | 21-07-2023 | Statistics Assignment : Questions based on Applied Statistics working with Mobile Phone dataset. | |

Table 21: Tasks for Week 5

| | |
|--------------------------------|-----------------|
| Name of the Supervisor: | Saiprasad Kagne |
| Supervisor's Sign: | |
| Date: | 22-07-2023 |
| Remarks by Supervisor: | |

Table 22: Supervisor's Remarks for Week 5

Weekly Summary - Week 5

Duration: From: 17-07-2023 To: 22-07-2023

Work Assigned:

| |
|------------------------|
| Statistics Assignment. |
|------------------------|

Table 23: Work assigned for Week 5

Learning Outcomes:

| |
|---|
| Deep understanding of concepts related to measures of central tendency. |
| Practical application of descriptive statistics concepts using an Employees dataset. |
| Application of statistics principles to analyze and interpret a Mobile Phone dataset. |

Table 24: Learning Outcomes for Week 5

Remarks:

| |
|---|
| Successfully applied measure of central tendency concepts to a practical scenario. |
| Analyzed and interpreted descriptive statistics using a real-world Employees dataset. |
| Demonstrated practical skills in applying statistics to analyze a Mobile Phone dataset. |
| Developed the ability to draw meaningful conclusions from statistical analysis in diverse contexts. |

Table 25: Remarks for Week 5

Name of the Supervisor: SaiPrasad Kagne

Sign of the Supervisor:

Weekly Report - Week 6

| S.No. | Date | Description of Work | Sign of Supervisor |
|-------|------------|--|--------------------|
| 1 | 24-07-2023 | Getting started with Tableau : Introduction, Tableau Options, Installation, Tableau Public and etc. | |
| 2 | 25-07-2023 | Business Challenge : Working with Datasets, Tableau Interface, Connecting Data and etc. | |
| 3 | 26-07-2023 | Creating Charts in Tableau : Charting Methods, Line Chart, Types of Bar Chart, Scatter Chart, Tree-map, Text Tables, Pie Chart and etc. | |
| 4 | 27-07-2023 | Charts Standout Techniques : Formatting Layout & Axes, Dual Axis, Combined Axis and Tool Tips. | |
| 5 | 28-07-2023 | Filters, Groups & Sets : Types of Filters, Filtering Measures, Groups & Sets Implementation. | |

Table 26: Tasks for Week 6

| | |
|--------------------------------|-----------------|
| Name of the Supervisor: | Saiprasad Kagne |
| Supervisor's Sign: | |
| Date: | 29-07-2023 |
| Remarks by Supervisor: | |

Table 27: Supervisor's Remarks for Week 6

Weekly Summary - Week 6

Duration: From: 24-07-2023 To: 29-07-2023

Work Assigned:

| |
|---|
| Create three diverse Tableau visualizations using a small dataset, apply filters, implement grouping or sets, and enhance visualizations with tooltips. |
|---|

Table 28: Work assigned for Week 6

Learning Outcomes:

| |
|---|
| Proficient Tableau usage from installation to navigating Tableau Public. |
| Mastery of diverse charting methods and advanced visualization techniques. |
| Effective application of filters, groups, and sets for refined data representation. |

Table 29: Learning Outcomes for Week 6

Remarks:

| |
|---|
| Successfully installed and utilized Tableau options, including Tableau Public. |
| Demonstrated practical dataset skills and chart creation mastery in Tableau. |
| Applied advanced charting techniques and refined data representation using filters, groups, and sets. |

Table 30: Remarks for Week 6

Name of the Supervisor: SaiPrasad Kagne

Sign of the Supervisor:

Weekly Report - Week 7

| S.No. | Date | Description of Work | Sign of Supervisor |
|-------|-------------------------|--|--------------------|
| 1 | 31-07-2023 | Calculated Fields & Parameters : Number Calculations, Logical Functions and Parameters. | |
| 2 | 01-08-2023 | Dashboards and Stories : Dashboard Layout, Action Filter, Highlight Action, Story and its Navigation Style. | |
| 3 | 02-08-2023 & 03-08-2023 | Tableau Case Study : Hands-on experience working with Tableau with the help of sample dataset. | |
| 4 | 04-08-2023 | Tableau Assignment : Major Dashboard Project using all above techniques at one place. | |

Table 31: Tasks for Week 7

| | |
|--------------------------------|-----------------|
| Name of the Supervisor: | Saiprasad Kagne |
| Supervisor's Sign: | |
| Date: | 05-08-2023 |
| Remarks by Supervisor: | |

Table 32: Supervisor's Remarks for Week 7

Weekly Summary - Week 7

Duration: From: 31-07-2023 To: 05-08-2023

Work Assigned:

| |
|---|
| Create three diverse Tableau visualizations using a small dataset, apply filters, implement grouping or sets, and enhance visualizations with tooltips. |
|---|

Table 33: Work assigned for Week 7

Learning Outcomes:

| |
|--|
| Proficient use of calculated fields, logical functions, and parameters in Tableau. |
| Mastered dashboard creation, layout design, and interactive actions. |
| Practical Tableau experience through a hands-on case study with a sample dataset. |
| Application of all techniques in a major Tableau assignment, creating a comprehensive dashboard project. |

Table 34: Learning Outcomes for Week 7

Remarks:

| |
|---|
| Integrated calculated fields, logical functions, and parameters in Tableau. |
| Achieved mastery in dashboard creation, layout design, and interactive actions. |
| Applied learned skills in a practical case study with a sample dataset. |
| Demonstrated proficiency in a major Tableau assignment, showcasing advanced dashboard project skills. |

Table 35: Remarks for Week 7

Name of the Supervisor: SaiPrasad Kagne

Sign of the Supervisor:

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

In the dynamic landscape of today's digital era, data has emerged as the lifeblood of decision-making across various domains. The field of Data Science plays a pivotal role in extracting meaningful insights from the vast volumes of data generated daily. This report aims to shed light on the importance of data analytics in diverse domains, emphasizing its transformative power in guiding critical decisions. Through the exploration of two distinct projects – one focusing on Heart Disease Diagnostic Analysis and the other delving into Swiggy Data Analysis – this report seeks to underscore the broad applications of data science, ranging from healthcare to the customer behavior in burgeoning realm of online food delivery.

1.2 SCOPE

Data analytics has become a cornerstone in driving informed decision-making processes, offering unprecedented opportunities to unlock patterns, trends, and correlations within data sets. The projects presented herein embody the quintessence of data science applications, illustrating its efficacy in addressing challenges across industries. The healthcare sector benefits from data-driven insights to enhance diagnostic capabilities, while the realm of online food delivery relies on analytics for optimizing operations and customer experiences.

1.3 PURPOSE / OBJECTIVE OF THE PROJECTS

The objective behind this report is to spotlight the transformative potential of data analytics. With the ability to discern hidden patterns, predict future trends, and derive actionable intelligence, data science has transcended its niche to become an indispensable tool across sectors. The projects outlined – Heart Disease Diagnostic Analysis and Swiggy Data Analysis – exemplify the versatility of data science, showcasing its relevance in both life-saving medical applications and the thriving digital marketplace.

1.4 PROBLEM STATEMENT

1.4.1 Heart Disease Diagnostic Analysis:

Motivation: Cardiovascular diseases remain a leading cause of global morbidity and mortality. Timely and accurate diagnosis is crucial for effective intervention and treatment.

Problem Statement: Develop a robust data analytics model capable of predicting the likelihood of heart disease based on a set of clinical parameters. Utilize historical patient data to create valuable insights to make data-driven decisions that assists healthcare professionals in early detection and risk stratification.

1.4.2 Swiggy Data Analysis:

Motivation: In the competitive landscape of the online food delivery industry, optimizing operational efficiency and enhancing customer satisfaction are paramount for sustained success.

Problem Statement: Conduct a comprehensive analysis of Swiggy's data to identify key trends, customer preferences, and operational bottlenecks. Propose data-driven strategies to enhance delivery efficiency, personalize user experiences, and improve overall service quality.

1.5 LIMITATIONS

1.5.1 Heart Disease Diagnostic Analysis:

1. Data Quality and Availability: The effectiveness of the analysis heavily relies on the quality and quantity of available data. Incomplete or inaccurate patient records could compromise the accuracy of analysis.

2. Ethical and Privacy Concerns: The utilization of sensitive medical data raises ethical and privacy concerns. Strict adherence to data protection regulations is essential, and obtaining consent from individuals for using their health information may pose challenges.

3.Generalization to Diverse Populations: Analysis made from specific demographic data may not generalize well to diverse populations. Variations in healthcare practices, genetic factors, and lifestyle choices can limit the insights and derived decision's applicability across different regions or communities.

4. Dynamic Nature of Health Data: Health data is dynamic, with new parameters and diagnostic criteria evolving over time. The analysis may become outdated as medical practices and standards progress, requiring regular updates to maintain accuracy.

1.5.2 Swiggy Data Analysis:

1. Data Accuracy and Completeness: Inaccuracies or missing data in the datasets, such as incomplete order information or customer preferences, can hinder the accuracy of the analysis and compromise the reliability of the insights generated.

2. Limited Access to Internal Data: Certain critical data, especially related to Swiggy's internal operations, may be proprietary or not available for analysis. This limitation can restrict a comprehensive understanding of the entire service chain.

3. External Factors Impacting Operations: External factors such as weather conditions, traffic, or unforeseen events can significantly impact delivery times and operational efficiency. Incorporating real-time external data into the analysis may be challenging.

4. User Behavior Variability: User preferences and behaviors are subject to change, influenced by various factors like trends, seasons, or marketing promotions. Predicting long-term user preferences accurately poses a challenge due to this inherent variability.

5. Competitive Landscape: External market dynamics, including the strategies and innovations of competitors, can affect Swiggy's performance. Analyzing these external factors comprehensively may be challenging, limiting the depth of strategic recommendations.

CHAPTER 2

LITERATURE SURVEY

2.1 Paper-1: Predictions in heart disease using techniques of data mining by Monika Gandhi, Computer Science and Engineering Department ASET, Amity University, Noida, India; Shailendra Narayan Singh, Amity University, Noida, Uttar Pradesh, India[1]

Abstract:- As huge amount of information is produced in medical associations (healing facilities, therapeutic focuses) yet this information is not properly utilized. The health care system is "data rich" however "knowledge poor ". There is an absence of successful analysis methods to find connections and patterns in health care data. Data mining methods can help as remedy in this circumstance. For this reason, different data mining techniques can be utilized. The paper intends to give details about various techniques of knowledge abstraction by using data mining methods that are being used in today's research for prediction of heart disease. In this paper, data mining methods namely, Naive Bayes, Neural network, Decision tree algorithm are analyzed on medical data sets using algorithms.

2.2 Paper-2: Using data mining techniques in heart disease diagnosis and treatment by Mai Shouman; Tim Turner; Rob Stocker, School of Engineering and Information Technology, Australian Defence Force Academy, University of New South Wales, Canberra, ACT, Australia[2]

Abstract:- The availability of huge amounts of medical data leads to the need for powerful data analysis tools to extract useful knowledge. Researchers have long been concerned with applying statistical and data mining tools to improve data analysis on large data sets. Disease diagnosis is one of the applications where data mining tools are proving successful results. Heart disease is the leading cause of death all over the world in the past ten years. Several researchers are using statistical and data mining tools to help health care professionals in the diagnosis of heart disease. Using single data mining technique in the diagnosis of heart disease has been comprehensively investigated showing acceptable levels of accuracy. Recently,

researchers have been investigating the effect of hybridizing more than one technique showing enhanced results in the diagnosis of heart disease. However, using data mining techniques to identify a suitable treatment for heart disease patients has received less attention. This paper identifies gaps in the research on heart disease diagnosis and treatment and proposes a model to systematically close those gaps to discover if applying data mining techniques to heart disease treatment data can provide as reliable performance as that achieved in diagnosing heart disease.

2.3 Paper-3: An Analysis of Online Food Ordering Applications in India: Zomato and Swiggy, Anupriya Saxena, Amity University, ABS, Lucknow, Uttar Pradesh, India[3]

Abstract:- Startups in India become a talk of the town in world business scenario. Youngsters are coming up with innovative concepts to counterpart untouched concerned area of consumers. Currently Indian Online food market is \$350 billion. Food technology in broad area, online food delivery apps are just part of it. This conceptual study will give more insight about emerging innovative technologies in restaurant industry and strategies followed by online food start ups Zomato, Swiggy. From this research paper we would understand drivers of online food sites. Different services given by application that makes consumers happy and satisfied. Comfort and Convenience which makes consumer more inclined towards online food ordering. The research concluded that due to urbanization in Indian landscape, online food delivery applications are growing with flying colors. Future of online food ordering website is bright. Facilities, Comfort, User friendliness are the key features of online websites success.

2.4 Paper-4: Food Delivery Services and Customer Preference: A Comparative Analysis by Natarajan Chandrasekhar, Saloni Gupta & Namrata Nanda[4]

Abstract:- This paper has revealed that the consumer perception plays a crucial role in understanding the decision-making process of the consumers. The purpose of the study was to determine the impact of online food delivery services like Swiggy, Foodpanda, Zomato, etc., on consumers. The research was entirely on primary data. A structured questionnaire was prepared, which had four parts, in other words, preference, reliability, consistency, and preference decision. The sample size was 169. The response rate was 84.5%. The data analysis technique used was grey analysis technique, which was used to interpret the result of the data collected. The

result showed that the consumers mostly prefer uniqueness in terms of price, quality, and delivery. None of the online food delivery services like Zomato, Swiggy or Foodpanda was given the highest rank. This study has given an overall view about what the consumers think, what are the uncertainties that they are facing, whether the service is reliable or not. It has given complete information about the problems that the consumers are facing and how it can be improved. From a managerial perspective, the paper contributes to understanding the consumers in a broader way. Based on the study of consumer perception, it has helped in examining various factors like preference, reliability, liking, etc. related to consumers adding novelty to this research.

CHAPTER 3

ANALYSIS

3.1 INTRODUCTION

The analysis of implementation is a critical phase in any data analysis project, as it focuses on translating the insights and findings derived from data into practical and actionable outcomes. This phase bridges the gap between data exploration and the application of results to address specific objectives or solve real-world problems.

3.2 REQUIREMENTS

3.2.1 FUNCTIONAL REQUIREMENTS

1. **Data Preprocessing:** Implement data cleaning, transformation, and normalization processes to ensure the dataset is suitable for analysis.
2. **Descriptive Statistical Analysis:** Perform descriptive statistical analysis to understand the distribution, central tendency, and variability of key variables.
3. **Data Visualization:** Create insightful visualizations to present trends, patterns, and relationships within the data.
4. **Correlation and Trends Analysis:** Explore correlations and trends within the data to uncover potential insights.

3.2.2 NON-FUNCTIONAL REQUIREMENTS

1. **Performance:** Ensure timely execution of data analysis tasks.
2. **Scalability:** Support the scalability of the system to handle growing datasets.
3. **Reliability:** Ensure the system's reliability in terms of handling errors and interruptions during analysis.

3.2.3 SOFTWARE REQUIREMENTS

1. **Programming languages:** Python
2. **Python Environment:** Python 3.x installed with necessary libraries (Pandas, NumPy, Matplotlib, Seaborn) for data manipulation and analysis.
3. **Data Analysis and Visualization libraries:** NumPy, Pandas, Matplotlib
4. **Tableau Desktop:** Tableau Desktop for creating interactive visualizations and dashboards.

5. Integrated Development Environment (IDE): Choose a Python IDE like Jupyter Notebook or Visual Studio Code or Google Colab for writing and executing Python scripts.

3.2.4 HARDWARE REQUIREMENTS

Processor: A multicore processor (e.g., Quad-core) for parallel processing.

RAM: Sufficient RAM, especially for handling large datasets. A minimum of 8GB is recommended, and higher for more extensive analyses.

Internet Connectivity: A stable internet connection is necessary for accessing online resources, installing packages, and displaying outputs.

3.3 FLOWCHART

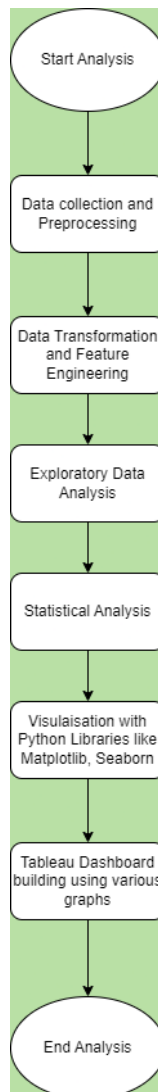


Figure 3.1: Flowchart of the Data Analysis and Visualization Process

CHAPTER 4

SYSTEM DESIGN OR METHODOLOGY

4.1 INTRODUCTION

In the realm of data-driven decision-making, the design of a system is the linchpin that transforms raw data into actionable insights. This document provides an in-depth exploration of the system design underpinning our data analysis projects, seamlessly merging the power of Python libraries — Matplotlib, NumPy, Pandas, and Seaborn—with the dynamic visualization capabilities of Tableau. This approach envisions a structured framework where data flows seamlessly from collection and preprocessing to sophisticated analytics and finally materializes in an interactive Tableau dashboard. This system design narrative not only unravels the technical intricacies but also underscores the pivotal role it plays in shaping the narrative of our data, empowering stakeholders with the tools needed to make informed decisions in an era increasingly defined by the language of data.

4.2 DATA FLOW DIAGRAM

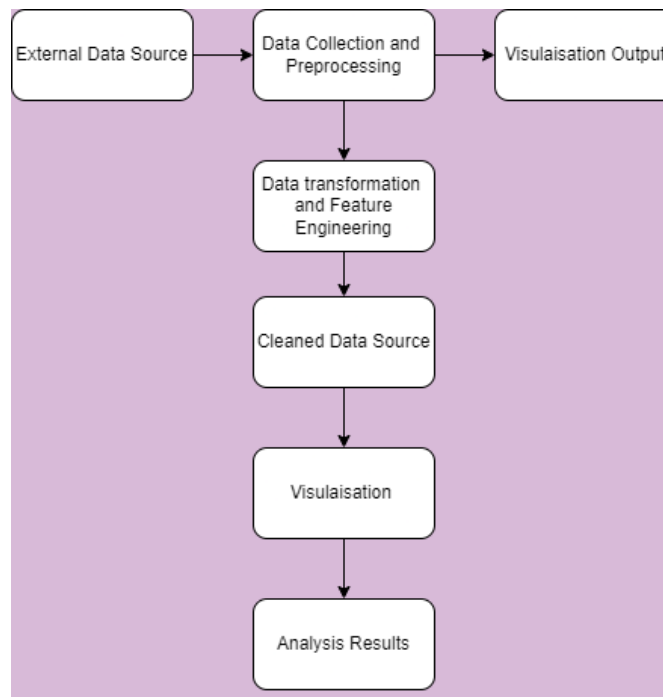


Figure 4.1: DFD

4.3 MODULE DESIGN AND ORGANSIATION

4.3.1 Heart Disease Diagnostic Analysis

1. Objective: Analyze data related to heart disease and develop insights to aid in diagnostic analysis.

2. Data Source: Medical records, possibly including patient demographics, vital signs, and diagnostic results.

The dataset considered has the following details:

age

sex

chest pain type (4 values)

resting blood pressure

serum cholesterol in mg/dl

fasting blood sugar >120 mg/dl

resting electrocardiographic results (values 0,1,2)

maximum heart rate achieved

exercise induced angina

oldpeak = ST depression induced by exercise relative to rest

the slope of the peak exercise ST segment

number of major vessels (0-3) colored by flourosopy

thal: 0 = normal; 1 = fixed defect; 2 = reversible defect

| A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|-----|-----|----|----------|------|-----|---------|---------|-------|---------|-------|----|------|--------|
| age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
| 52 | 1 | 0 | 125 | 212 | 0 | 1 | 168 | 0 | 1 | 2 | 2 | 3 | 0 |
| 53 | 1 | 0 | 140 | 203 | 1 | 0 | 155 | 1 | 3.1 | 0 | 0 | 3 | 0 |
| 70 | 1 | 0 | 145 | 174 | 0 | 1 | 125 | 1 | 2.6 | 0 | 0 | 3 | 0 |
| 61 | 1 | 0 | 148 | 203 | 0 | 1 | 161 | 0 | 0 | 2 | 1 | 3 | 0 |
| 62 | 0 | 0 | 138 | 294 | 1 | 1 | 106 | 0 | 1.9 | 1 | 3 | 2 | 0 |
| 58 | 0 | 0 | 100 | 248 | 0 | 0 | 122 | 0 | 1 | 1 | 0 | 2 | 1 |
| 58 | 1 | 0 | 114 | 318 | 0 | 2 | 140 | 0 | 4.4 | 0 | 3 | 1 | 0 |
| 55 | 1 | 0 | 160 | 289 | 0 | 0 | 145 | 1 | 0.8 | 1 | 1 | 3 | 0 |
| 46 | 1 | 0 | 120 | 249 | 0 | 0 | 144 | 0 | 0.8 | 2 | 0 | 3 | 0 |
| 54 | 1 | 0 | 122 | 286 | 0 | 0 | 116 | 1 | 3.2 | 1 | 2 | 2 | 0 |
| 71 | 0 | 0 | 112 | 149 | 0 | 1 | 125 | 0 | 1.6 | 1 | 0 | 2 | 1 |
| 43 | 0 | 0 | 132 | 341 | 1 | 0 | 136 | 1 | 3 | 1 | 0 | 3 | 0 |
| 34 | 0 | 1 | 118 | 210 | 0 | 1 | 192 | 0 | 0.7 | 2 | 0 | 2 | 1 |
| 51 | 1 | 0 | 140 | 298 | 0 | 1 | 122 | 1 | 4.2 | 1 | 3 | 3 | 0 |
| 52 | 1 | 0 | 128 | 204 | 1 | 1 | 156 | 1 | 1 | 1 | 0 | 0 | 0 |
| 34 | 0 | 1 | 118 | 210 | 0 | 1 | 192 | 0 | 0.7 | 2 | 0 | 2 | 1 |
| 51 | 0 | 2 | 140 | 308 | 0 | 0 | 142 | 0 | 1.5 | 2 | 1 | 2 | 1 |
| 54 | 1 | 0 | 124 | 266 | 0 | 0 | 109 | 1 | 2.2 | 1 | 1 | 3 | 0 |
| 50 | 0 | 1 | 120 | 244 | 0 | 1 | 162 | 0 | 1.1 | 2 | 0 | 2 | 1 |
| 58 | 1 | 2 | 140 | 211 | 1 | 0 | 165 | 0 | 0 | 2 | 0 | 2 | 1 |
| 60 | 1 | 2 | 140 | 185 | 0 | 0 | 155 | 0 | 3 | 1 | 0 | 2 | 0 |
| 67 | 0 | 0 | 106 | 223 | 0 | 1 | 142 | 0 | 0.3 | 2 | 2 | 2 | 1 |
| 45 | 1 | 0 | 104 | 208 | 0 | 0 | 148 | 1 | 3 | 1 | 0 | 2 | 1 |
| 63 | 0 | 2 | 135 | 252 | 0 | 0 | 172 | 0 | 0 | 2 | 0 | 2 | 1 |
| 42 | 0 | 2 | 120 | 209 | 0 | 1 | 173 | 0 | 0 | 1 | 0 | 2 | 1 |
| 61 | 0 | 0 | 145 | 307 | 0 | 0 | 146 | 1 | 1 | 1 | 0 | 3 | 0 |

Figure 4.2: Heart Disease Diagnosis Dataset

3. Python Libraries Used:

- Matplotlib: for data visualization.
- NumPy: for numerical operations.

- Pandas: for data manipulation and analysis.
- Seaborn: for statistical data visualization.

4. Analysis Steps:

1. Data Cleaning: Handle missing values, outliers, and inconsistencies.
2. Exploratory Data Analysis (EDA): Analyze trends, patterns, and correlations in the data.
3. Statistical Analysis: Apply statistical methods to draw meaningful conclusions.
4. Visualization: Use Matplotlib and Seaborn to create visualizations for better understanding.
5. Tableau Dashboard Design: Create a Tableau dashboard to combine insights from the analysis for a comprehensive view on filters for demographic details in the heart disease analysis.

5. Outcome: Insights into factors influencing heart disease, visualizations for better understanding, and potentially a diagnostic model.

4.3.2 Swiggy Data Analysis

1. Objective: Analyze data related to Swiggy, a food delivery platform, to extract actionable insights.

2. Data Source: Swiggy Bangalore Outlet data of restaurant locations, ratings, cuisines, and average costs, etc.,

The dataset considered has the following details:

ShopName

Cuisine

Location

Rating

Cost_for_two

| Shop_Name | Cuisine | Location | Rating | Cost_for_Two |
|--|---|------------------------------------|--------|--------------|
| Kanti Sweets | Sweets | Koramangala, Koramangala | 4.3 | 150 |
| Mumbai Tiffin | North Indian, Home Food, Thalys, Coml | Sector 5, HSR | 4.4 | 400 |
| Sri Krishna sagar | South Indian, North Indian, Fast Food, | 6th Block, Koramangala | 4.1 | 126 |
| Al Daaz | American, Arabian, Chinese, Desserts, | HSR, HSR | 4.4 | 400 |
| Beijing Bites | Chinese, Thai | 5th Block, Koramangala | 4.1 | 450 |
| Kitchens of Punjab | North Indian | Koramangala 4th Block, Koramangala | 4.2 | 350 |
| 99 VARIETY DOSA AND PAV BHAIJ- Malli Mane Food Court | Fast Food, North Indian, Chinese | BTM 2nd Stage, BTM | 4.1 | 200 |
| La Pino'z Pizza | Italian | BTM, BTM | 3.9 | 500 |
| Hotel Manu | South Indian, Kerala, Chinese, North In | HSR, HSR | 4.1 | 350 |
| Yumlane Pizza | Pizzas, Italian, Mexican | 9th Main road, Koramangala | 3.8 | 150 |
| Ambur Star Briyani | Chinese, South Indian, North Indian, D | outer ring road, BTM | 4.1 | 500 |
| Cake Box | Desserts | Koramangala, Koramangala | 4 | 247 |
| Meghana Foods | Chinese, Andhra, Biryani, Seafood | 5th Block, Koramangala | 4.3 | 550 |
| Momoz | Chinese | 5th Block, Koramangala | 4.3 | 450 |
| A2B - Adyar Ananda Bhavan | South Indian, Chinese, Desserts, North | 7th Block, Koramangala | 4.2 | 450 |
| Shawarma Inc | Arabian, Fast Food | 1st MAin, Koramangala | 4.1 | 150 |
| WarmOven Cake & Desserts | Desserts, Beverages | Koramangala, Koramangala | 4.1 | 200 |
| Sri Lakshmi Dhaba | North Indian | Bommanahalli, BTM | 3.7 | 200 |
| Falahaar & Kota Kachori | North Indian | 6th block, Koramangala | 4.2 | 300 |
| Shree Khana Khazana | Indian, Rajasthani | Sector 4, HSR | 4.1 | 350 |
| Just Bake - Cakes & confectioners | Desserts, Bakery | BTM 1st stage, BTM | 4.3 | 300 |
| Maa Di Hatti | Chinese, Healthy Food, North Indian | Jakkasandra Extn, Koramangala | 4 | 129 |
| Hotel Godavari | North Indian, Chinese, Hyderabad | Marutnagar Main Road, BTM | 4 | 400 |
| Rolls On Wheels | Fast Food | 1st Block, Koramangala | 4.1 | 300 |
| New Udupi Grand | Chinese, Jain, North Indian, South Indi | HSR, HSR | 4.3 | 150 |

Figure 4.3: Swiggy-Bangalore Outlet Dataset

3. Python Libraries Used:

- Matplotlib: for data visualization.
- NumPy: for numerical operations.
- Pandas: for data manipulation and analysis.
- Seaborn: for statistical data visualization.

4. Analysis Steps:

1. Data Cleaning: Handle missing values, outliers, and inconsistencies.
2. Exploratory Data Analysis (EDA): Analyze trends, patterns, and correlations in the data.
3. Statistical Analysis: Apply statistical methods to draw meaningful conclusions.
4. Customer Feedback Analysis: Explore reviews and feedback data to understand customer sentiments.
5. Visualization: Use Matplotlib and Seaborn to create visualizations for better understanding.
6. Tableau Dashboard Design: Create a Tableau dashboard to combine insights from the analysis for a comprehensive view on filters for demographic details in the heart disease analysis.
7. Business Recommendations: Provide actionable insights for improving service and customer satisfaction.

5. Outcome: Insights into customer behavior, visualizations for decision-makers, and recommendations for business improvement.

CHAPTER 5

IMPLEMENTATION

The implementation of the project is done as follows:

5.1 HEART DISEASE DIAGNOSTIC DATA ANALYSIS

5.1.1 Data Understanding and Preparation:

- Collecting the "Heart Disease Diagnostic Data" and understanding its structure.
- Cleaning and preprocessing the data, handling missing values and outliers using various functions.
- Exploring basic statistics and visualizing the distribution of key features.

5.1.2 Exploratory Data Analysis (EDA):

- Conducting EDA to identify patterns, correlations, and trends in the data.
- Using libraries like NumPy, Pandas.

5.1.3 Feature Engineering:

- Identifying relevant features and potentially engineer new ones like the age group and birth year from age.
- Normalizing or scaling features if needed.

5.1.4 Visualisation

- Drawing various bar plots, histograms, correlation heatmaps, piecharts, to analyse or find relations between attributes
- Use visualisations to understand the data and draw meaningful insights
- Using visualisation tools and libraries in Python like Matplotlib, and Seaborn.

5.1.5 Tableau Dashboard Creation:

- Importing the cleaned and preprocessed data into Tableau.
- Designing a dashboard with relevant visualizations such as:
 1. *A Stacked Bar chart rate of heart diseases based on various(4) chest pain types among men and women.*
 2. *A Scatter Plot of Maximum heart Rate and ST depression.*
 3. *A Bar graph showing heart disease rate among men and women.*

4. A Pie Chart showing proportion of individuals with exercise induced angina.

5. A Stacked Bar chart showing distribution of Men and Women with Heart Diseases across age groups.

- Using filters and parameters to make the dashboard interactive.

5.1.6 Interpretation and Insights:

- Interpreting the analysis results to understand which features contribute most to heart disease prediction.

- Determining which attribute is affecting the heart disease rate and among which genders or age groups is the risk of having a heart disease higher based on various medical parameters considered.

5.2 SWIGGY- BANGALORE OUTLET DATA ANALYSIS

5.2.1 Data Collection and Cleaning:

- Collecting the "Swiggy Data" containing information on orders, cuisines, locations, restaurants, ratings and costs.

- Cleaning the data, handling any missing or inconsistent values using inbuilt functions and libraries available in Python.

5.2.2 Exploratory Data Analysis (EDA):

- Exploring customer behavior, order patterns, and ratings.
- Using Python libraries such as NumPy and Pandas.

5.2.3 Order Analysis:

- Identifying popular delivery areas and potential areas for expansion.
- Understanding what kind of cuisines are more ordered in order to include them more in restaurants.

5.2.4 Customer Segmentation:

- Segmenting customers based on ordering behavior, frequency, and preferences.
- Creating visualizations like cuisine clusters or personas.

5.2.5 Visualisation

- Drawing various bar plots, histograms, correlation heatmaps, piecharts, word clouds, scatter plots to analyse or find relations between attributes

- Use visualisations to understand the data and draw meaningful insights
- Using visualisation tools and libraries in Python like Matplotlib, and Sebaorn.

5.2.6 Tableau Dashboard Creation:

- Importing the cleaned and preprocessed data into Tableau.
- Designing a dashboard with relevant visualizations:
 1. *A Bar chart showing average costs per two spent in different restaurants and hotels in Kormanagala.*
 2. *A Bar graph showing number of shops per cuisine.*
 3. *A Tree map showing Ratings of shops in BTM.*
 4. *A Pie chart showing average cost for two in HSR.*
 5. *A Bar chart showing average costs per two spent in different restaurants and hotels in Kormanagala.*
- Using filters and parameters to make the dashboard interactive.

5.2.7 Insights and Presentation:

- Deriving actionable insights from the visualizations.
- Presenting findings to stakeholders using the Tableau dashboard.
- Highlighting trends, areas for improvement, and opportunities for business growth.

CHAPTER 6

RESULTS

6.1 HEART DISEASE DIAGNOSTIC ANALYSIS

The following are some of the different graphs obtained and insights developed:

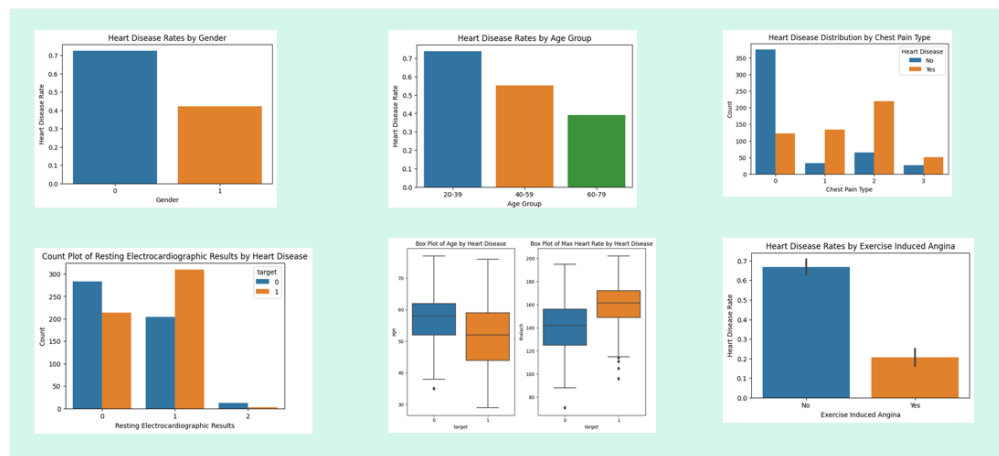


Figure 6.1: Visualizations created for Heart Disease Diagnostic Analysis

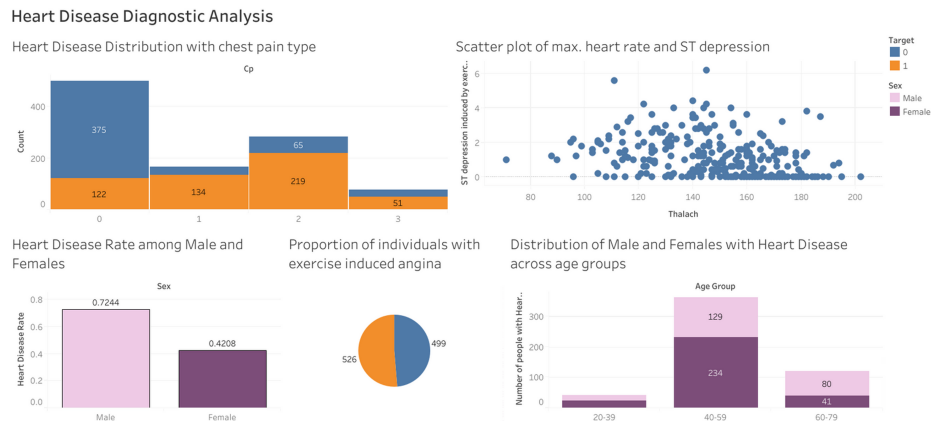


Figure 6.2: Tableau Dashboard created for the Heart Disease Diagnostic Analysis

6.2 SWIGGY - BANGALORE OUTLET DATA ANALYSIS

The following are some of the different graphs obtained and insights developed:

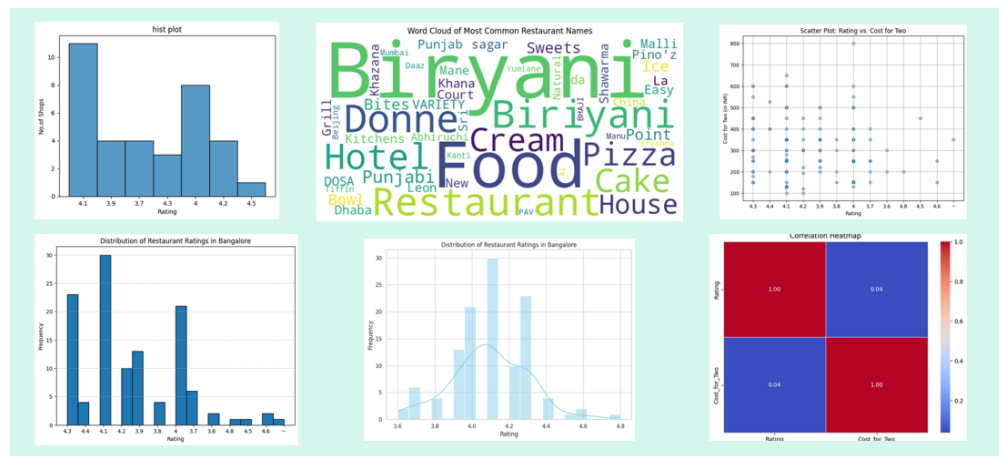


Figure 6.3: Visualizations created for Swiggy Data Analysis

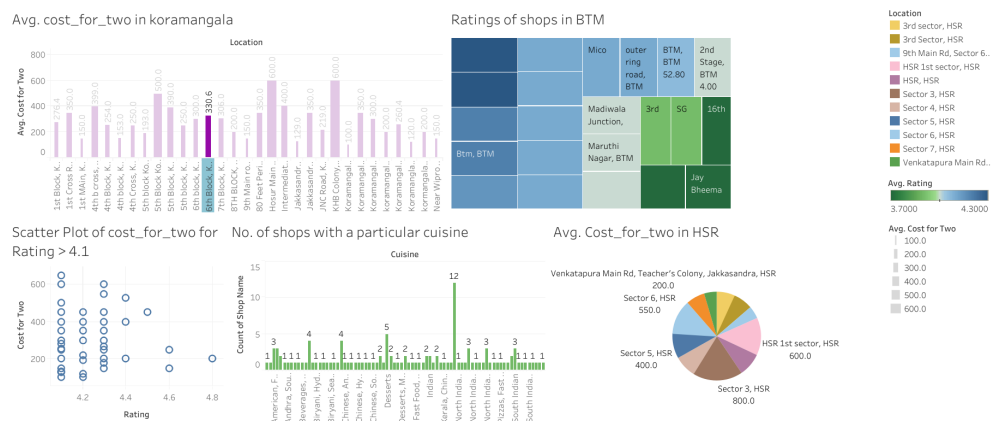


Figure 6.4: Tableau Dashboard created for the Swiggy Data Analysis

CHAPTER 7

FUTURE SCOPE

7.1 HEART DISEASE DIAGNOSTIC ANALYSIS

1. Personalized Medicine: Utilize advanced analytics, including machine learning and genetic data, to develop personalized treatment plans for individuals based on their unique health profiles.

2. Real-time Health Monitoring: Integrate wearable devices and IoT (Internet of Things) technology to monitor health parameters in real-time, providing early warnings and proactive interventions for individuals at risk of heart disease.

3. Explainable AI (XAI): Enhance the interpretability of machine learning models to gain deeper insights into the factors influencing heart disease predictions. This could improve trust in AI-driven diagnostic tools.

4. Longitudinal Studies: Conduct long-term studies by analyzing data over extended periods, allowing for a better understanding of the dynamic nature of heart disease risk factors and outcomes.

5. Integration with Electronic Health Records (EHR): Collaborate with healthcare providers to integrate analysis tools with electronic health records, facilitating seamless information flow and improving patient care.

7.2 SWIGGY DATA ANALYSIS

1. Dynamic Pricing and Promotions: Incorporate user reviews and sentiment analysis to gain insights into how specific mobile phone features influence price perception and overall user satisfaction.

2. Supply Chain Optimization: Use data analysis to optimize the supply chain, reducing delivery times, and improving overall efficiency. Predict demand patterns to manage inventory effectively.

3. Enhanced Customer Experience: Leverage natural language processing (NLP) and sentiment analysis to understand customer reviews and feedback, leading to improvements in service quality and customer experience.

4. Predictive Analytics for Demand: Develop predictive models to anticipate demand for specific cuisines, enabling better resource allocation and operational planning.

5. Integration with Emerging Technologies: Explore integration with emerging

technologies such as augmented reality (AR) for interactive menus or blockchain for secure and transparent transactions.

6. Sustainability and Environmental Impact: Conduct analyses on the environmental impact of food delivery services and implement sustainable practices. This could include optimizing delivery routes to reduce carbon footprint.

7. Market Expansion Strategies: Analyze data to identify potential markets for expansion, considering factors such as demographics, cultural preferences, and economic indicators.

CHAPTER 8

CONCLUSION

In conclusion, data analysis projects exemplified by "Heart Disease Diagnostic Analysis" and "Swiggy Data Analysis" stand at the forefront of transformative endeavors, showcasing the remarkable potential of leveraging data for informed decision-making and improved outcomes. The "Heart Disease Diagnostic Analysis" not only enables early detection and prevention of cardiovascular risks but also empowers healthcare professionals with nuanced insights, optimizing resource allocation and contributing to ongoing research. Simultaneously, "Swiggy Data Analysis" redefines the landscape of food delivery services, offering enhanced customer experiences through personalized recommendations and operational efficiency. These projects underscore the multidimensional benefits of data analysis, ranging from early disease intervention and resource optimization in healthcare to adaptive marketing strategies and sustainability initiatives in the food delivery industry. The fusion of technology, analytics, and domain-specific knowledge not only refines existing processes but also propels innovation, setting the stage for a future where data-driven insights play a central role in shaping healthcare practices and business strategies. As we navigate this data-driven era, the enduring impact of these projects highlights the imperative of harnessing the power of data to foster positive change and drive progress across diverse domains.

BIBLIOGRAPHY

- [1] Monika Gandhi; Shailendra Narayan Singh. “Predictions in heart disease using techniques of data mining”. In: *International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)* (2015).
- [2] Mai Shouman; Tim Turner; Rob Stocker. “Using data mining techniques in heart disease diagnosis and treatment”. In: *Japan-Egypt Conference on Electronics, Communications and Computers* (2012).
- [3] Lucknow Uttar Pradesh India Anupriya Saxena Amity University ABS. “An Analysis of Online Food Ordering Applications in India: Zomato and Swiggy”. In: *4th National Conference On Recent Trends in Humanities, Technology, Management Social Development (RTHTMS 2K19)* 9 (April 2019).
- [4] Saloni Gupta Namrata Nanda Natarajan Chandrasekhar. “Food Delivery Services and Customer Preference: A Comparative Analysis”. In: *Journal of Food-service Business Research* 22 (Jun 2019).