MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

Final Year Internship Report

on

Data Science Intern at Tiger Analytics

Submitted By:

Karishma Parashar

0901CS181053

Faculty Mentor:

Dr. Ranjeet Kumar Singh

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957

MAY-JUNE 2022

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

Data Science Intern at Tiger Analytics

A final year internship report submitted in partial fulfilment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by:

Karishma Parashar

0901CS181053

Internship Faculty Mentor:

Dr. Ranjeet Kumar Singh

Submitted to:

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957

MAY-JUNE 2022

Internship Certificate Received from Industry/Company

Karishma Parashar
Jan 18, 2022

Dear Karishma,

We are pleased to extend to you an offer of internship with **Tiger Analytics India Consulting Private Limited** (the **Company**).

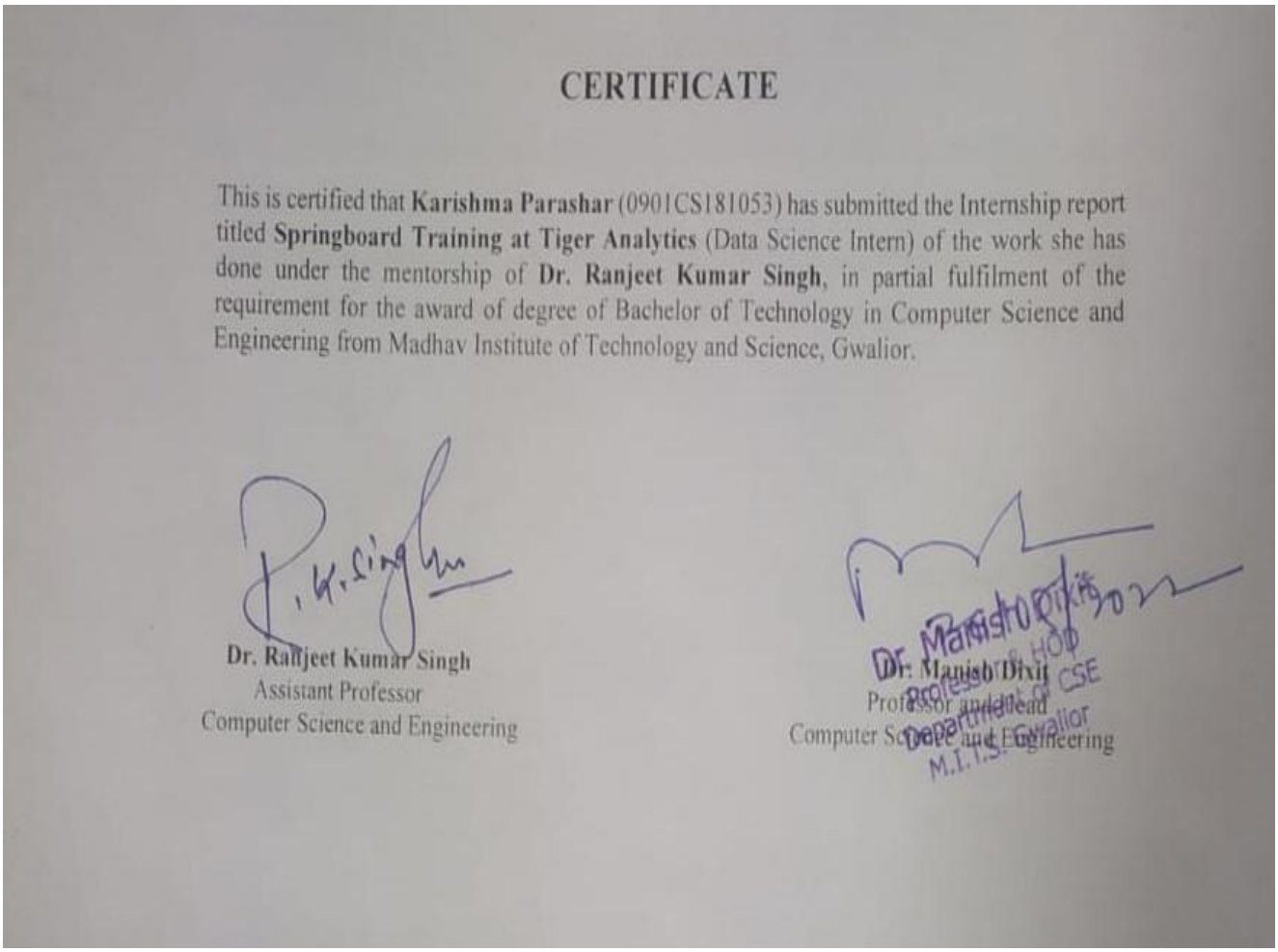
This contract is valid only for the period of internship, and you will be required to sign a separate contract should you take up a full-time role with the Company.

Your internship is subject to the following terms and conditions:

1. Date of Commencement
   The internship is for a period of 4 Months - Jan 24, 2022 to May 31, 2022

Since internship is still ongoing, certificate is not available as of now, therefore, as a proof, I've attached snippet of the offer letter for now. However, I'll submit the certificate at the time of no dues.

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)



CERTIFICATE

This is certified that **Karishma Parashar** (0901CS181053) has submitted the Internship report titled **Springboard Training at Tiger Analytics** (Data Science Intern) of the work she has done under the mentorship of **Dr. Ranjeet Kumar Singh**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering from Madhav Institute of Technology and Science, Gwalior.

Dr. Ranjeet Kumar Singh
Assistant Professor
Computer Science and Engineering

Dr. Manish Dixit
Professor and Head
Computer Science and Engineering

Dr. Manish Dixit
Professor & HOD
Department of CSE
M.I.T.S. Gwalior

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

DECLARATION

I hereby declare that the work being presented in this Internship report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in CSE at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of Dr. Ranjeet Kumar Singh, Assistant Professor, Department of CSE.

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Karishma Parashar
0901CS181053
IV Year,
Computer Science and Engineering

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

ACKNOWLEDGEMENT

The full semester internship has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary internship as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr.Manjaree Pandit** for this.

I would sincerely like to thank my department, **Department of Computer Science and Engineering, for allowing** me to explore this internship. I humbly thank **Dr. Manish Dixit**, Professor and Head, Department of Computer Science and Engineering, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Dr Ranjeet Kumar Singh.**, Assistant Professor, Department of Computer Science and Engineering, for his continued support and close mentoring throughout the internship. I am also very thankful to the faculty and staff of the department.

Karishma Parashar
0901CS181053
IV Year,
Computer Science and Engineering

ABSTRACT

This Data Science Internship at Tiger Analytics, aims at making the candidate job ready by instilling all kinds of necessary skills required to become a successful data analyst. Not just the training programs but a live project mentored by an establishedpersonnelinthisdomain.DataAnalyticsdealswithhandling,cleaning,pre-processing, interpreting, analysing, and making meaningful inferences out of the data.This internship commenced on 24-Jan-2022 and is still ongoing. This will finish offafter results of 8thsemester exams, thus converting my role of an intern into full timeDataAnalystat TigerAnalytic

TABLE OF CONTENTS

TITLE	PAGE NO.
Internship Certificate from Industry	iii
Institute Internship Certificate	iv
Declaration	v
Acknowledgement	vi
Abstract	vii
List of Figures	2
List of Abbreviation	3
Chapter 1: Introduction	4-8
Chapter 2: Requirement Analysis	8-10
Chapter 3: System Requirement Specifications	11
Chapter 4: Technologies used during internship	12
Chapter 5: Work done during internship	12-22
Chapter 6: Applications	22-23
Chapter 7: Conclusion	23
References	23
Appendices	24-30

LIST OF Figures

Fig.No.	Fig.Caption	Page No.
1.1	Company Logo	4
3.1	Hardware Specification	11
5.1	Project Problem Image	13
5.2	Process Flow Diagram	14
5.3	Tools image	15
5.4	Distribution of Ratings	17
5.5	Approxcost_of_two Vs Rating Graph	18
5.6	Affordable Graph	18
5.7	Top 15 Cheapest Restaurants Graph	19
5.8	Top 15 Expensive Restaurants Graph	19
5.9	Distribution of Cuisine Graph	20
5.10	Cuisine Overall Analysis	20
5.11	Cuisine Analysis-BTM Area	21
5.12	Cuisine Analysis-HSR Area	21
5.13	Cuisine Analysis-Koramangala Area	22
6.1	Application of DataScience	23

LIST OF ABBREVIATIONS

S.No.	Abbreviation	Description
1.	MS	Microsoft
2.	SQL	Structured Query Language
3.	DBMS	Database Management System
4.	OOPS	Object Oriented ProgrammingSystem
5.	ML	Machine Learning
6	MAPE	Mean Absolute Percentage Error
7.	VIF	Variance Inflation Factor

Chapter 1: INTRODUCTION

1.1 About Company:

Fig1.1

Tiger Analytics is pioneering what AI and analytics can do to solve some of the toughest problems faced by organizations globally. We develop bespoke solutions powered by data and technology for several Fortune500 companies. We have offices in multiple cities across the US, UK, India, and Singapore, and a substantial remote global workforce. We have received multiple awards ranging from being recognized as a Leader by Forrester Research to being ranked among the fastest-growing tech companies by Inc. and Financial Times. We consistently feature in prestigious 'Best Analytics Firms' lists. If you're interested to explore career opportunities with Tiger, we'd love to tell you more. You'll work alongside the best in the industry, learn the structured approach to innovate, and have fun along the way.

1.2 About CEO

Mahesh Kumar is the Founder and CEO of Tiger Analytics. He started Tiger Analytics with a desire to bring his experience in management science to help organizations achieve superior performance through the application of advanced analytics. Before founding Tiger Analytics, Mahesh was on the faculty of the Smith School of Business and Rutgers Business School.

He has conducted research in the areas of data mining and statistical modeling and has successfully applied his research to solve problems related to forecasting, pricing, promotions, and customer segmentation for a wide range of businesses across various verticals. Mahesh holds a Ph.D. in Operations Research and Marketing from MIT, and a B.Tech in Computer Science from IIT Bombay

1.3 Modules

1.3.1 MS Excel

The value of MS Excel has increased significantly in the business environment. MS Excel is used in Data Science for data storage purposes as well as data management and data detection. Not everyone is familiar with technical jargon, so MS Excel works as easy to read and has extensive acceptance and support. For someone with a basic knowledge of MS Excel it is very important that they live in the corporate world, whether it is an entry level job or a GenMan role. The topics taught were:

1: Applications of Office

2: Types of excel sheets

3: Different data types

4: Tables

5: Graphs

6: Data Cleaning

7: Data handling

8: Data Manipulation

9: Data Visualization

10: Formulae implementation

We learnt in Excel

- How we can analyse our dataset
- How different types of graphs can be made and make conclusion from them

1.3.2 SQL

SQL stands for Structured query Language. SQL is an easy-to-learn language designed specifically to work with information sites. There is a growing need for professionals who can manage information. Almost all large companies use SQL. It is widely used in various fields such as ticket booking, banking, social media, data sharing, e-Commerce, etc. SQL plays its role in functions such as data analyst, data engineer, webmaster, etc. and, again, is an important skill you can have to be successful as a data analyst. The lessons taught are:

1: What is Database

2: Basics of DBMS

3: My SQL workbench.

4: Database design

5: Normalization

6: SQL basic commands

7: SQL advanced commands

8: Functions

9: Merging operations

We learnt in SQL

- How to extract our desired information from tables.
- How to merge different tables and then extract data according to our need.

1.3.3 Python Programming

Python is a translated, dynamic typed language with precise and efficient syntax. Also, Python can be easily integrated with other programming languages. A complete course of python was taught over Udemy in order to achieve fluency in python and familiarity with the application of python in domain of Data Science. The topics taught were:

1: Python Crash Course

2: Python for Data Analysis- NumPy

3: Python for Data Analysis- Pandas

4: Python for Data Visualization- Matplotlib

5: Python for Data Visualization- Seaborn

We learnt in Python

- How to use different types of library present in python.
- How to use inbuilt function of numpy and pandas library for data manipulation
- How to use inbuilt function of matplotlib and seaborn for data visualization.

1.3.4 GIT and GitHub

Anyone working in IT sector is well aware of the role and importance of git in a developer's life. Being a necessity, it was taught along with GitHub, working on Linux in order to get wider support at various levels of work.

We learnt in Git And Github

- How to maintain our code at one place so our teammates and manager can access our code and can make changes whatever they want.

1.3.5 Statistical Concepts

Statistics is backbone of data science. As it is a common knowledge that data science deals with organisation, manipulation, and making inferences out of data, the inferential statistics comes into play. Statistics here is not the regular stats but computations of data using various testing methods, validations, amendments, conclusions and a lot more. Topics taught were:

1: Intro to Statistics

2: Sample vs Population

3: Descriptive Statistics

4: Measures of central tendency

5: Distributions

6: Estimators vs Estimates

7: Confidence Intervals

8: Inferential Statistics

9: Hypothesis Testing

We learnt in Statistical Concepts

- Statistical concepts are used to review , analyze and draw conclusions from data.

1.3.6 Regression Analysis

Regression models (both linear and non-linear) are used for predicting a real value, like salary for example. If your independent variable is time, then you are forecasting future values, otherwise your model is predicting present but unknown values.

The following Machine Learning Regression models:

1. Simple Linear Regression
2. Multiple Linear Regression
3. Polynomial Regression
4. Support Vector for Regression (SVR)
5. Decision Tree Regression

We learnt in Regression Analysis

- Linear regressions can be used in business to evaluate trends and make estimates or forecasts.

1.3.7 Classification Analysis

Unlike regression where you predict a continuous number, you use classification to predict a category. Classification models include linear models like Logistic Regression, SVM, and nonlinear ones like K-NN, Kernel SVM and Random Forests.
The following Machine Learning Classification models:

1. Logistic Regression
2. K-Nearest Neighbours (K-NN)
3. Support Vector Machine (SVM)
4. Kernel SVM
5. Naive Bayes
6. Decision Tree Classification
7. Random Forest Classification

Chapter 2: Requirement Analysis

For a project to be successful, it is very important to analyze the project needs as they are collected and throughout the life cycle of the project. Needs analysis helps keep needs in line with business needs. A good needs analysis process will provide a software program that addresses the business objectives set. Requirement Analysis is the process of defining what users expect from an application to be built or modified. Needs analysis involves all activities that are performed to identify the needs of different stakeholders. Needs analysis therefore means analyzing, writing, verifying and managing software or system requirements. The requirements for high quality are documented, feasible, measurable, scalable, traceable, helping to identify business opportunities, and are defined as simplifying system design. To understand, let's have a hypothetical case of classroom. N students having their marks across two subjects A and B.

Life cycle in a Data Science Project

2.1 Phase 1

Data Analysis: Data analysis is the process of cleaning, modifying, and processing raw data, as well as extracting relevant information that helps businesses make informed decisions. The process helps to reduce the risks involved in decision-making by providing useful information and statistics, often presented in charts, diagrams, tables, and graphs.

In Data Analysis , the following analysis is done-

1. Missing Values
2. All the Numerical Variables
3. Distribution of the Numerical Variables
4. Categorical Variables
5. Cardinality of Categorical Variables
6. Outliers
7. Relationship between independent and dependent feature

2.2 Phase 2

Feature Engineering: Feature engineering refers to the process of using domain knowledge to select and transform the most relevant variables from raw data. The goal of feature engineering and selection is to improve the performance of machine learning (ML) algorithms.

The following steps are performed in Feature Engineering:

1. Missing values
2. Temporal variables
3. Categorical variables: remove rare labels
4. Standardize the values of the variables to the same range

2.3 Phase 3

Feature Selection: Now after feature engineering , Feature Selection is done. In feature selection, we get rid of the irrelevant and redundant features from the original data to optimize the overall performance. It is one of the most important phase before model building.

2.4 Phase 4

Model Building: The following steps are performed during model building-

1. We split the dataset into train and test data . By default , the test data is 30 % of the actual data and the training data is 70 % of the actual data.
2. After splitting the data, x_train contains the independent variables of the training data and y_train contains the dependent variable of training data. Same is done for the test data.
3. Then we build the base model and check its performance based on these parameters – R^2value, MAPE, bias, variance etc.
4. Then we check for multicollinearity between the independent variables. If the degree of correlation is high, it can cause problem of overfitting.
5. To remove multicollinearity between the independent variables, VIF (Variance inflation factor) is used.
6. After removal of some features using VIF, again feature selection is done. If some of the features that are important in answering the business questions are removed, so manual selection for those features is done.

2.5 Phase 5

Model Deployment: Out of all the different models (for example – linear regression, random forest, decision tree etc) we try during model building, the model which gives the best results and answers all the business questions is deployed. This is the final phase in the life cycle of a data science project.

Chapter 3 : System Requirement Specifications

Following are the system requirements for a data analyst:

Hardware Specifications:

Resource	Requirements	
Operating systems (64-bit)	Microsoft Windows	• Windows Server 2008 R2 • Windows Server 2008 R2 Standard • Windows Server 2008 R2 Enterprise • Windows Datacenter 2008 R2 • Windows Server 2012 • Windows Server 2012 R2 • Windows 2012 R2 Datacenter • Windows Server 2016 • Windows Server 2016 Datacenter • Windows Server 2019
	Linux	• Red Hat Enterprise Linux versions 6.x, 7.x • SUSE Linux Enterprise Server 11.x, 12.x, 15 • Oracle Linux Server 6.x, 7.x
	The preceding list is applicable for the product and the target hosts where data resides that you might install on the target hosts.	
CPU and memory	2 CPUs and 4 GB RAM **Note:** In a production environment, the minimum requirement would be 8 CPUs and 16 GB RAM for handling a 50-100 GB volume of data per day and up to four users.	

Fig3.1

Software Requirements:

1. Jupyter Notebook
2. Visual Studio Code
3. Ubuntu
4. MS Excel
5. Git

Chapter 4: Technologies used during internship

Data Science comprises of a multitude of technologies. It consists of various roles, which work at various levels, incorporating various technologies, helping world manage their data in an effective and efficient way. Following are the technologies used in Data Science.

1: Python Programming
2: SQL
3: MS Excel
4: Git
5: Unix
6: Regression Analysis
7: Statistical Analysis

Chapter 5: Work done during internship

Work done during the internship is divided into assignments given in each module mentioned above and a project solved using different technologies that we learned and implemented during the internship.

5.1 Assignments

- Tiger Takeaway Assignment – Python
- Tiger Takeaway Assignment – Advanced Excel
- Tiger Takeaway Assignment – SQL
- Tiger Takeaway Assignment – MLE Training
- Practice Assignment – Stats Concepts
- Practice Assignment – Regression
- Practice Assignment – Classification

5.2Project:Swiggy Data Analysis

Fig5.1

Problem Statement: Food industries are having important reflection of the economy from past few decades. In this project, we are analyzing the various aspects with different use cases which covers many aspects of Swiggy Food Delivery Service. It helps in not only understanding the meaningful relationships between attributes, but it also allows us to do our own research and come-up with our findings.

Analysis of the following is performed in this Project:

1. Distribution of 'Rating'.
2. Area-wise Analysis in terms of 'Rating' and 'Cost_for_Two (₹)' on BTM, HSR, & Koramangala Area.
3. Analyse "Approx Cost of 2 People" vs "Rating". Find out the relationship between them.
4. Analyze Affordable/Budgeted and Highest Rated Restaurants of Bangalore.
5. Top 15 Cheapest & Highest Rated Restaurants with Approx. Cost for 2 People.
6. Top 15 Expensive & Highest Rated Restaurants with Approx. Cost for 2 People.
7. Most preferred Cuisines by the Customers.
8. Area-wise Cuisines Analysis & Distribution of Cuisines in BTM, HSR, & Koramangala (Bangalore) Restaurants.

Dataset:

Download the dataset for this project from following Link -

- https://drive.google.com/drive/folders/12kcRkn5NGDh78RfKsRyTfasbaQRV7STN

Data Analysis

In the datasets Following features (columns) are given:

1. Shop_Name: Name of the Shop/Restaurants.
2. Cuisine: Name of the different Cuisines provided by Restaurants.
3. Location: Restaurant Area/Location.
4. Rating: Rating given by the Customers out of 5.
5. Cost_for_Two (₹): Approx. Cost of Two people w.r.t. Restaurants.

Out of the 5 features 2 are Continuous and 3 are Categorical features.

Process-Flow

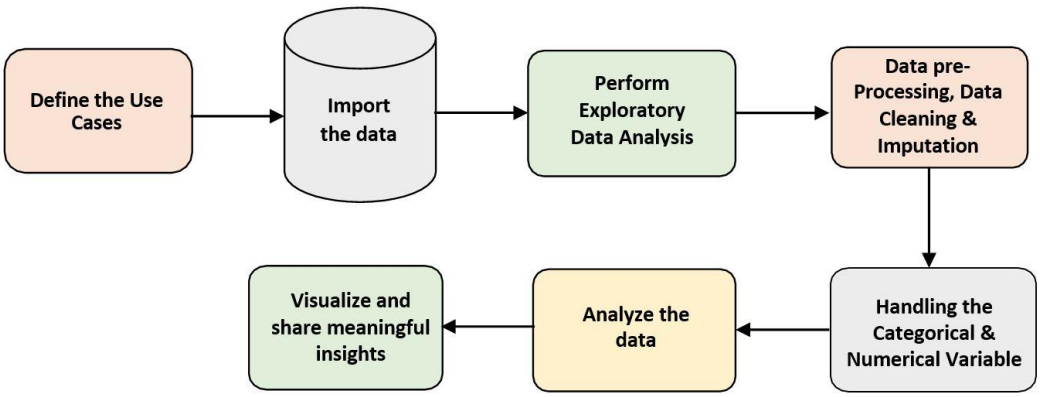


Fig5.2

Tools Used

- Jupyter Notebook is used as IDE.
- Numpy and Pandas is used for Data Manipulation & Pre-processing.
- For visualization of the plots, Matplotlib, Seaborn, Plotly are used.

Fig 5.3

Code: Download the dataset for this project from following Link –

- https://drive.google.com/drive/folders/12kcRkn5NGDh78RfKsRyTfasbaQRV7STN

Result: In this analysis project, we have been analysed several different use cases for the given dataset t o make better business decisions and help analyse customer trends and satisfaction, which can lead to new and better products and services.

✓ **Distribution of Rating**
- From this Distribution Plot, we can conclude that *More that '50%' of Restaurants* are having a *Rating* greater than "4.1" with a *Maximum Rating* of "4.8" which is considered as a decent Rating.
- And it also means that, most of these Restaurants are doing very well &rated accordingly by the Customers.

✓ **Area-wise Analysis in terms of 'Rating' and 'Cost_for_Two (₹)'**

• In BTM Area: Most of the Restaurants has 4.0 to 4.2 Rating and Approx. Cost for Two People lies between 200 to 350. (Max. Cost goes up to 600)
•HSR: Most has 4 or above Rating and Approx. Cost for Two People lies between 300 to 400. (Max. Cost goes up to 800)
•Koramangala: Most has 4.0 to 4.3 Rating and Approx. Cost for Two People lies between 200 to 350. (Max. Cost goes up to 600)

With this we can conclude the Costliest Area is HSR.

✓ **Analyse "Approx. Cost of 2 People" vs "Rating".**
- From this 'Scatter Plot', We can clearly say that - Bigger the Bubble, Higher the Price is... Similarly, Smaller the Bubble, Lesser the Price is...
- Furthermore, we can also say that *Most of the Affordable/Budgeted Restaurants* are having *Excellent Rating* as well.
- Same we can see, For *Approx. Cost* of "200", "150", "250", and "450", the *Ratings* were "4.8", "4.6", and "4.5" respectively.

- This might be because *Most of the people prefer Affordable/Budget-Restaurants* which also provides *good quality* of *Cuisines*.
- and on the other hand, there are *few Expensive Restaurants* who *doesn't* have that much *Rating* and they are *Expensive*.
- *Restaurants* which *Cost* around "600" to "800" for *Two People* are having the *Ratings in between '4.0' to '4.1'* which is *too less* as compared to *Affordable/Budgeted Restaurants*.

So, that's a Conclusion we can draw from this 'Scatter Plot'.

✓ **Analyse Affordable/Budgeted and Highest Rated Restaurants of Bangalore.**

- We have also analysed that, we have Total "82" which are the "Budget Restaurants" as well as they are "Affordable".
- On top of that, we have found out, Most of the Affordable/Budgeted Restaurants are having Excellent Rating as well. Like, For Approx. Cost of "200", "150", "250", and "450", the Ratings were "4.8", "4.6", and "4.5" respectively.
- This might be because Most of the people prefer Affordable/Budget-Restaurants which also provides good quality of Cuisines.
- And on the other hand, there are few Expensive Restaurants who doesn't have that much Rating and they are Expensive too.
- Those Restaurants Costs around "600" to "800" for Two People are having the Ratings in between '4.0' to '4.1' which is too less as compared to Affordable/Budgeted Restaurants.

✓ **Most preferred Cuisines by the Customers.**

- From the above Visualizations, we can say, *Most of the Restaurants* sell "Chinese" which is around '12.9%' followed by "North Indian" & "South Indian" Cuisines which are around '11.8%' & '8.46%'.

So, we can also infer that Most of the people are fond of these Cuisines.

✓ **Area-wise Cuisines Analysis & Distribution of Cuisines in Bangalore**

- In BTM Area, Most of the Restaurants sell "Chinese" which is around '17.1%' followed by "North Indian" & "South Indian" Cuisines which are around '15.2%' & '9.52%'. So, we can also infer that Most of the people are fond of these Cuisines.
- In HSR Area, "North Indian" Cuisines are dominated by around '14.3%' followed by "Chinese" & "South Indian" Cuisines '9.52%' & '9.52%' Restaurants respectively.
- In Koramangala Area, "Chinese" Cuisines are dominated by around '10.3%' followed by "North Indian" & "South Indian" Cuisines '9.66%' & '7.59%' Restaurants respectively.
-

Result Analysis

- Distribution of Rating

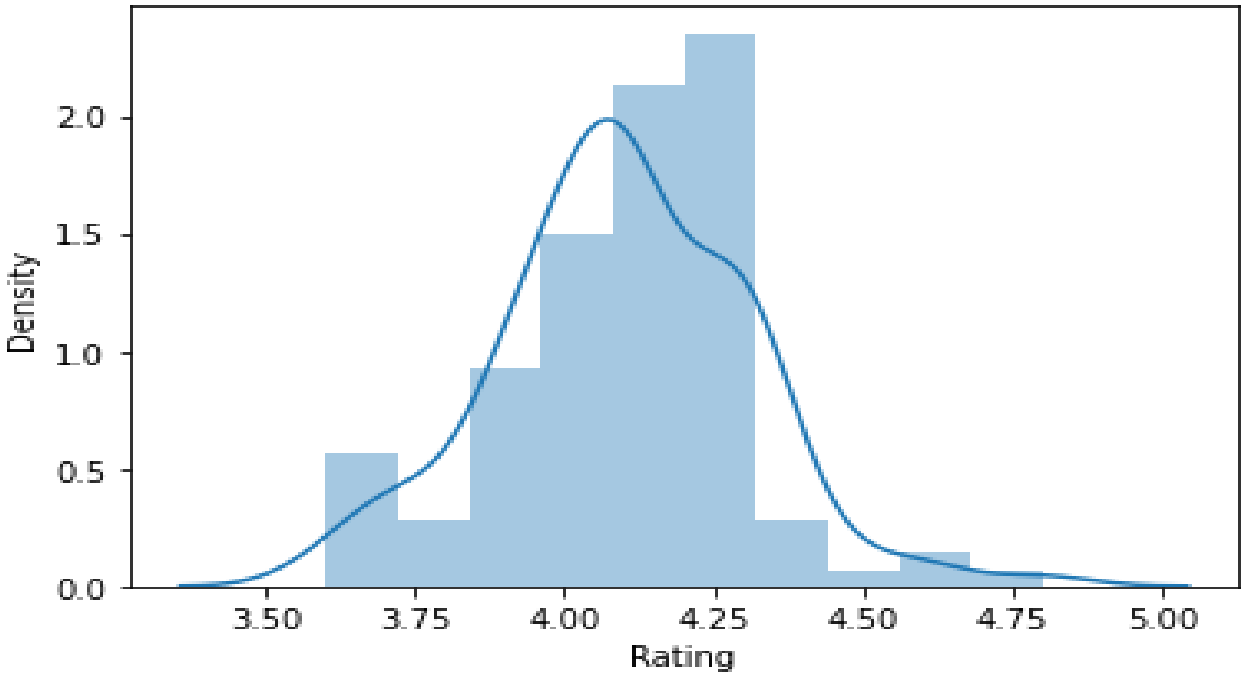


Fig5.4

- Analyse ‘approx. Cost of 2 people’ vs ‘Rating’

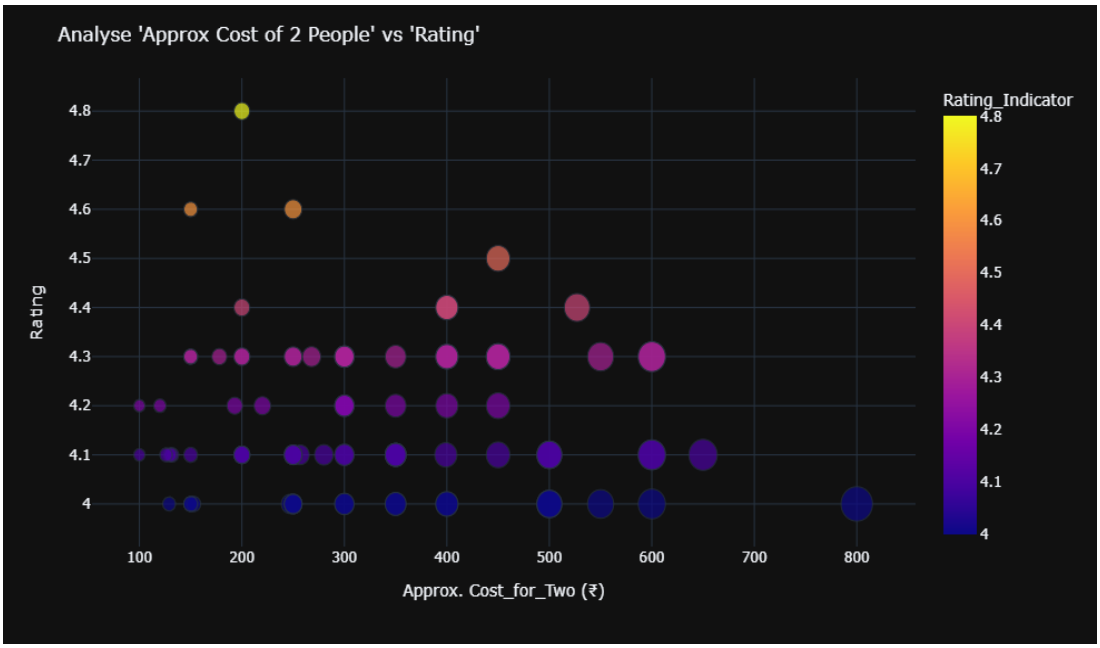


Fig5.5

- Affordable/Budgeted and Highest Rated Restaurants (Banglore)

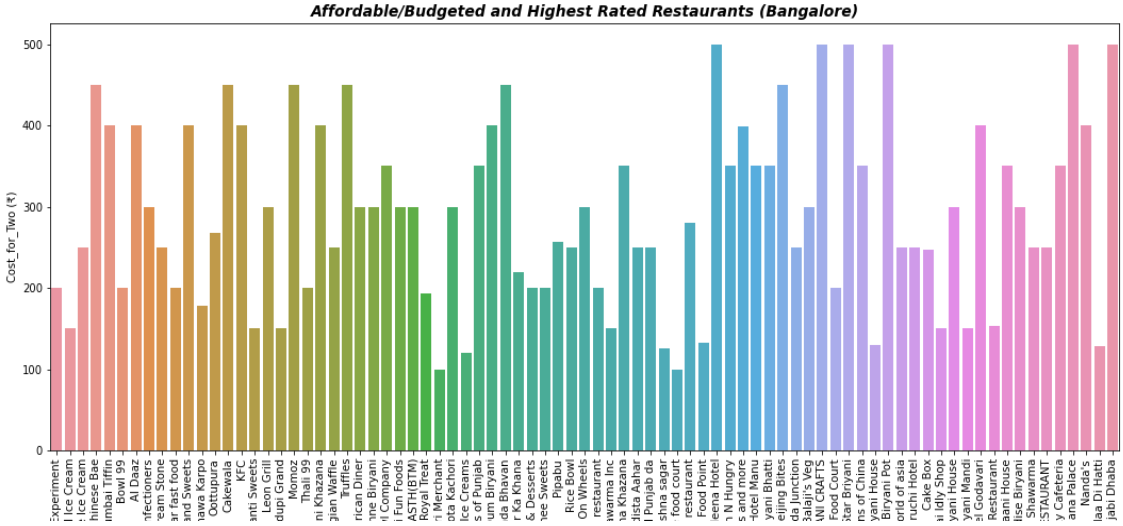


Fig5.6

- Top 15 Cheapest & Highest Rated Restaurants with Approx. Cost for 2 People.

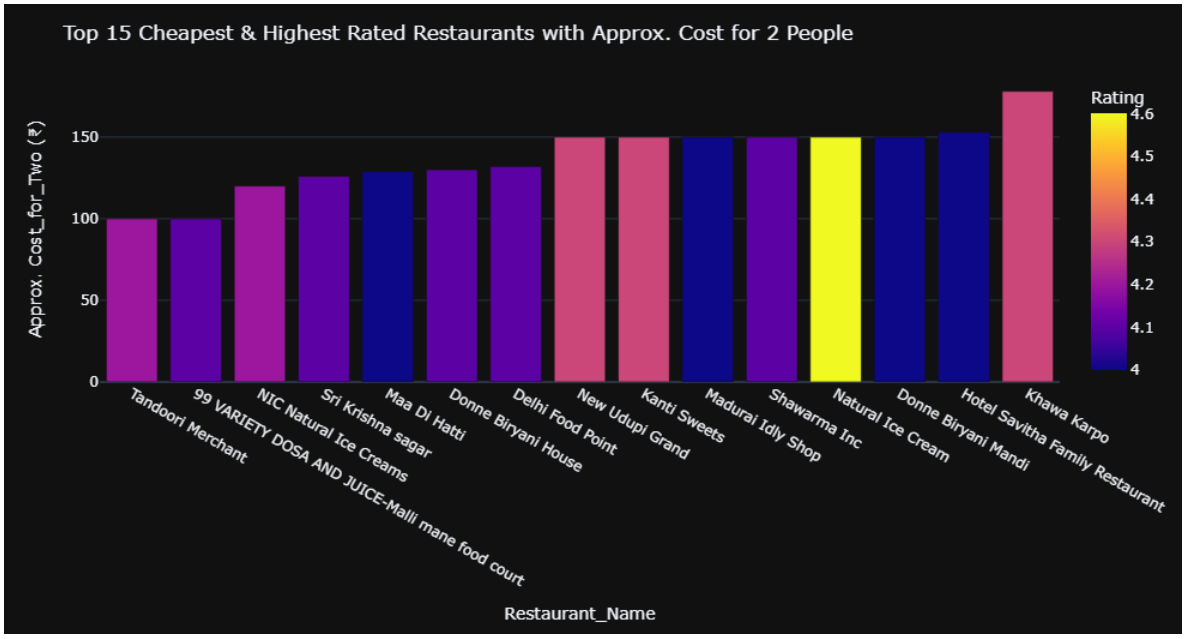


Fig5.7

- Top 15 Expensive& Highest Rated Restaurants with Approx. Cost for 2 People.

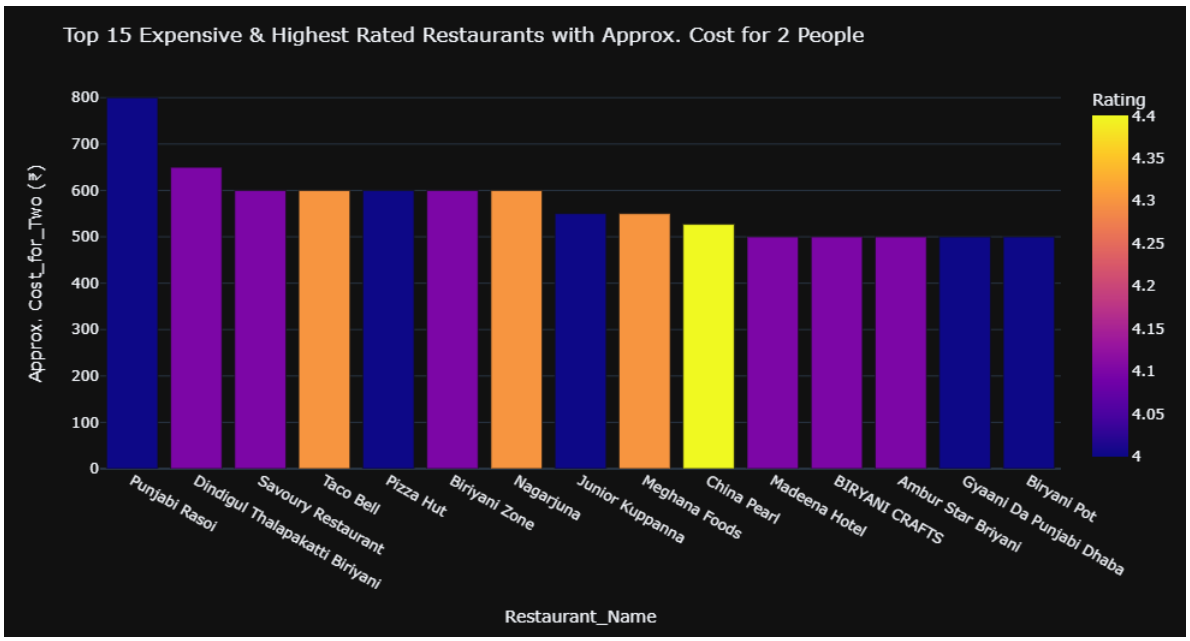


Fig5.8

- Distribution of Cuisines in Bangalore Restaurants

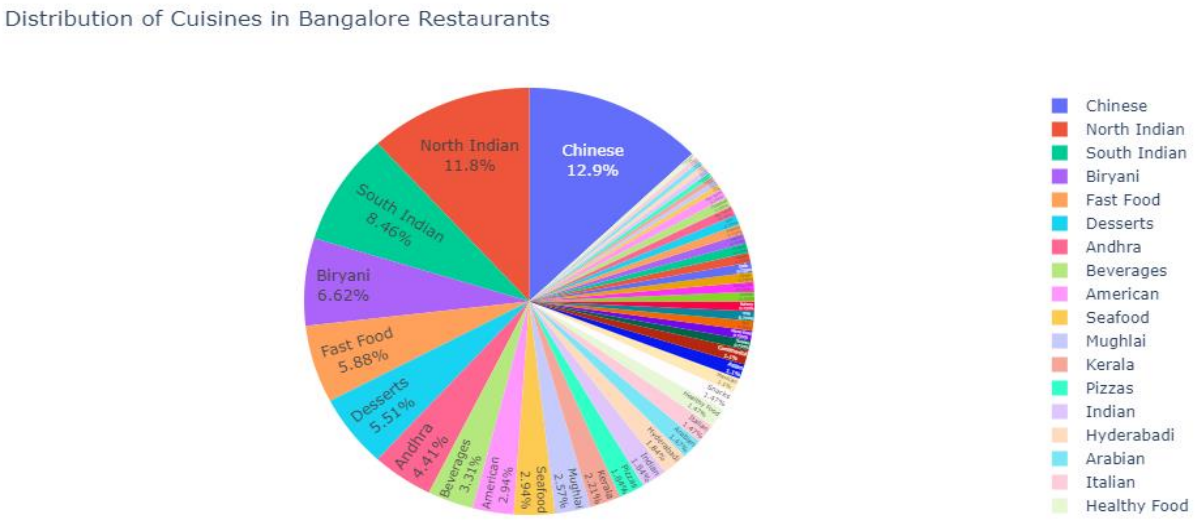


Fig5.9

- Cuisines Overall Analysis (Bangalore)

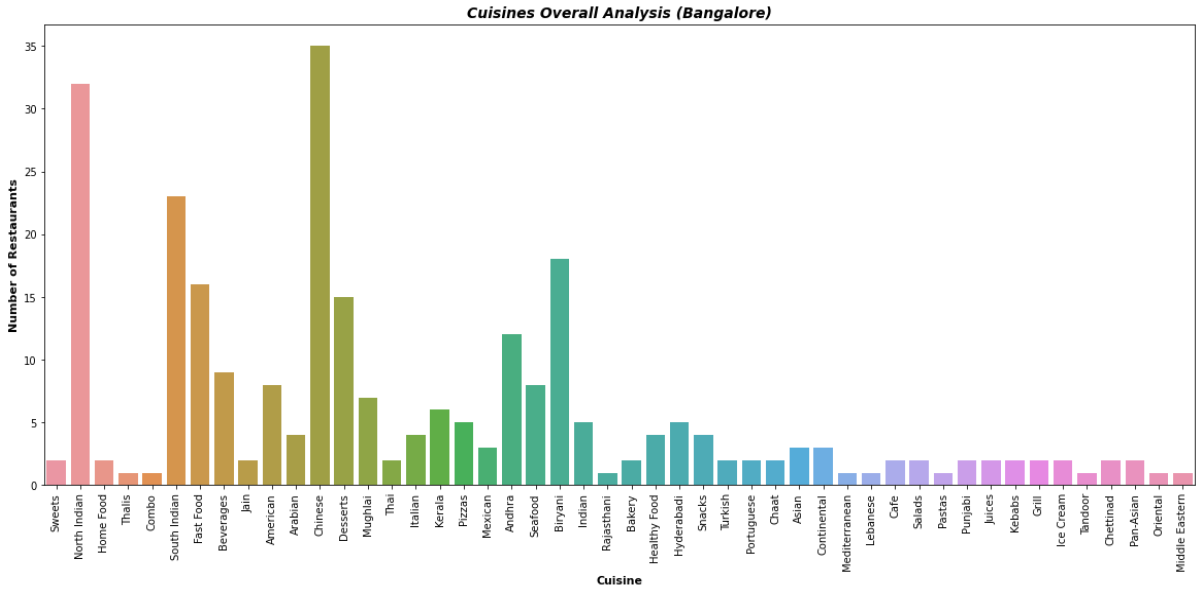


Fig5.10

- Cuisines Analysis- BTM (Bangalore)

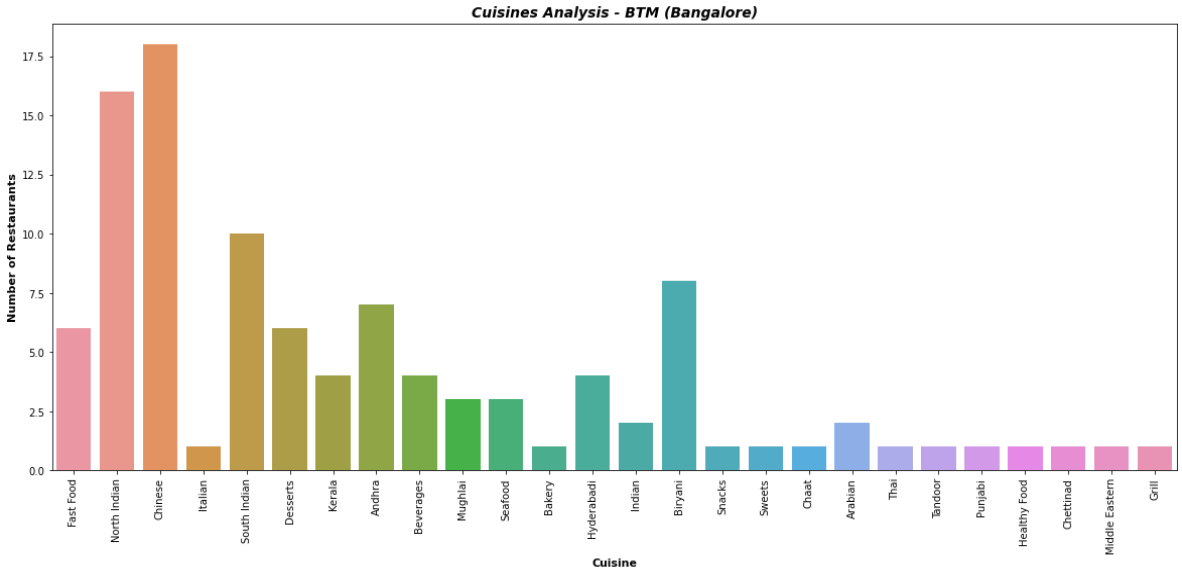


Fig5.11

- Cuisines Analysis- HSR (Bangalore)

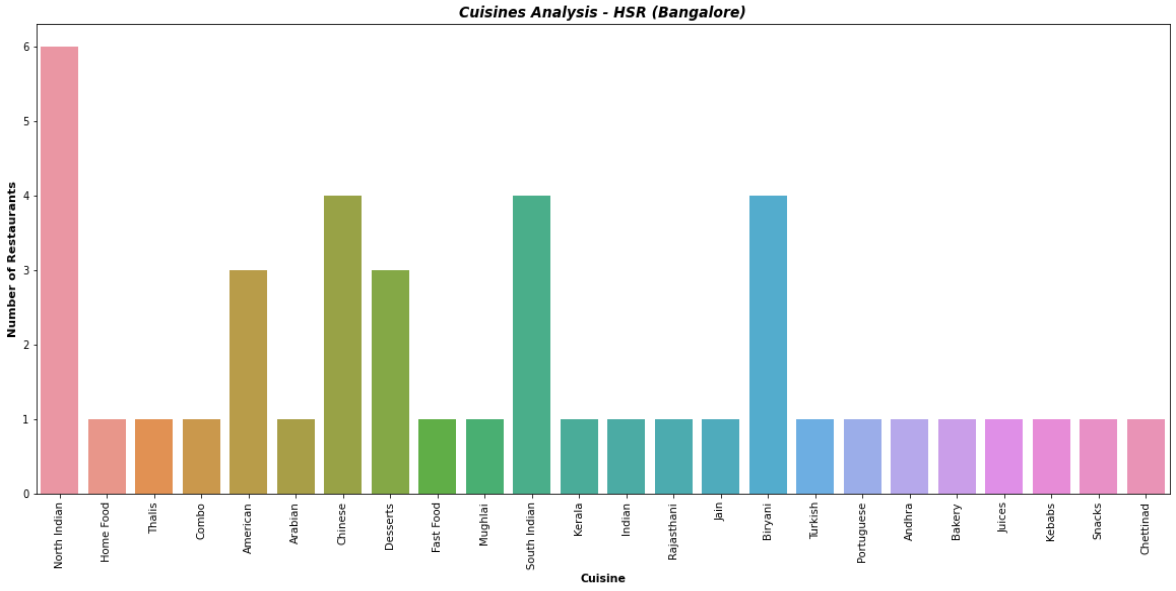


Fig5.12

- Cuisines Analysis- Koramangala (Bangalore)

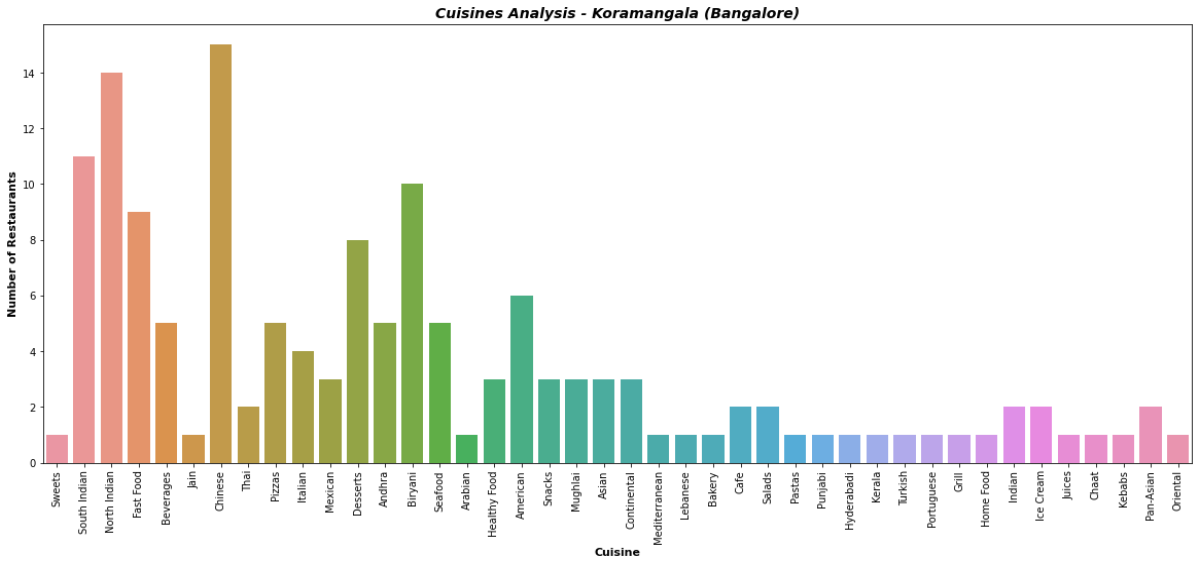


Fig5.13

Chapter 6: Applications

Data science is the field of applying advanced analytics techniques and scientific principles to extract valuable information from data for business decision-making, strategic planning and other uses.

Following are the applications of Data Science:

- To gain knowledge about behaviours and processes.
- Write algorithms that process large amounts of information quickly and efficiently.
- increase security.
- Privacy of sensitive data.
- Guide data-driven decision-making.
- Other applications

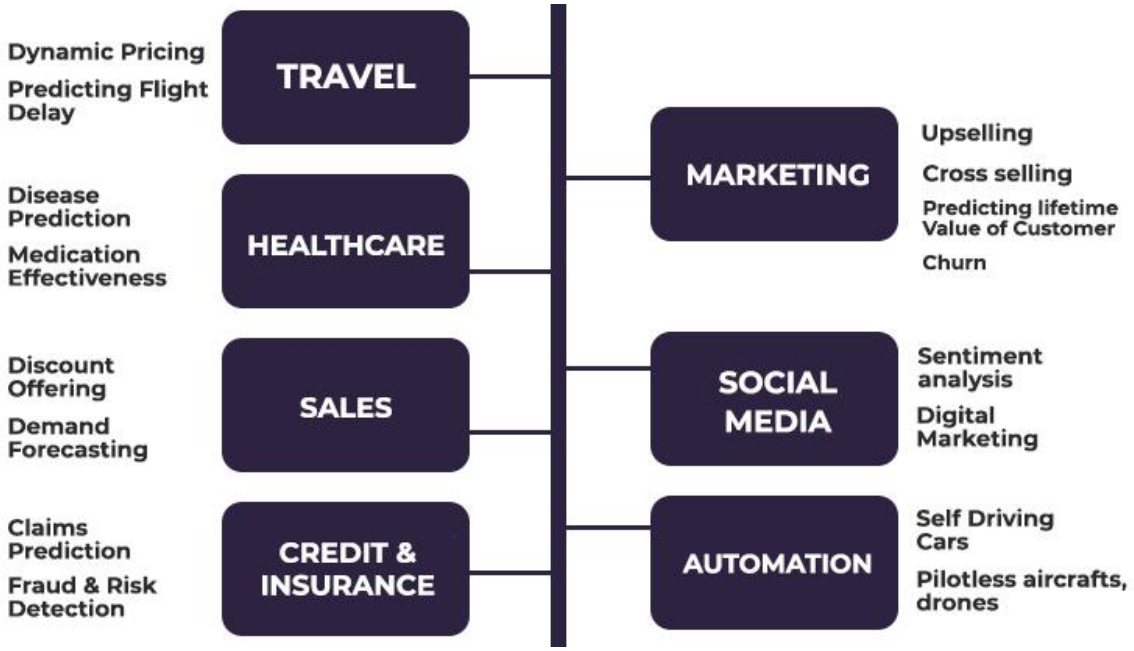


Fig6.1

Chapter 7: Conclusion

This Internship has been a really good phase of my life so far as not only on professional, but it has helped me a lot on personal front too. The key takeaways. I'd like to mention are:

1. I got exposure to corporate culture and real-world business.
2. Apart from relevant technologies, learnt a lot about team building, collaborative nature, and networking.
3. Learnt about delivery system in the domain of data science and how it is different from traditional IT company.
4. Learnt how to work and deliver under pressure and strict deadlines.

References

- Udemy provided by organization.
- https://www.statology.org/
- https://blog.minitab.com/en/adventures-in-statistics-2/how-to-interpret-regression-analysis-results-p-values-and-coefficients#:~:text=How%20Do%20I%20Interpret%20the,can%20reject%20the%20null%20hypothesis
- https://www.youtube.com/watch?v=ZftI2fEz0Fw&list=PLKnIA16_Rmvbr7zKYQuBfsVkjoLcJgxHH

Appendix for FPR

- First FPR

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Karishma Parashar		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	24 Jan-12 Feb	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work					
Learning capacity/Knowledge up gradation					
Performance/Quality of work					
Behaviour/Discipline/Team work					
Sincerity/Hard work					
Comment on nature of work done/Area/Topic	Karishma has started with the campus Training Program & started going through all the fundamental modules of data science. She is completing courses on time & performs well.				
OVERALL GRADE (Any one)	VERY GOOD				
Name of Industry Mentor	Padmajothi Murugaboopathy				
Signature of Industry Mentor	Padmajothi Murugaboopathy				

Receiving Date		**Name of Faculty Mentor**		**Sign**	

➢ Second FPR

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Karishma Parashar		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	13 Feb-3Mar	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work					
Learning capacity/Knowledge up gradation					
Performance/Quality of work					
Behaviour/Discipline/Team work					
Sincerity/Hard work					
Comment on nature of work done/Area/Topic	Karishma has started with the campus Training Program & started going through all the fundamental modules of data science. She is completing courses on time & performs well.				
OVERALL GRADE (Any one)	**VERY GOOD**				
Name of Industry Mentor	Padmajothi Murugaboopathy				
Signature of Industry Mentor	Padmajothi Murugaboopathy				

Receiving Date		**Name of Faculty Mentor**	P.K.Singh	**Sign**	P.K.Singh.

➢ Third FPR

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Karishma Parashar		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	1Mar-30Mar	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work					
Learning capacity/Knowledge up gradation					
Performance/Quality of work					
Behaviour/Discipline/Team work					
Sincerity/Hard work					
Comment on nature of work done/Area/Topic	Karishma has started with the campus Training Program & started going through all the fundamental modules of data science. She is completing courses on time & performs well.				
OVERALL GRADE (Any one)	**VERY GOOD**				
Name of Industry Mentor	Padmajothi Murugaboopathy				
Signature of Industry Mentor	Padmajothi Murugaboopathy				

Receiving Date		**Name of Faculty Mentor**	P.K.Singh.	**Sign**	P.K.Singh

➢ Fourth FPR

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Karishma Parashar		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	30Mar-13April	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work					
Learning capacity/Knowledge up gradation					
Performance/Quality of work					
Behaviour/Discipline/Team work					
Sincerity/Hard work					
Comment on nature of work done/Area/Topic	Karishma has completed fundamental modules of data science. Now working on Case Study She is completing courses on time & performs well.				
OVERALL GRADE (Any one)	**VERY GOOD**				
Name of Industry Mentor	Padmajothi Murugaboopathy				
Signature of Industry Mentor	Padmajothi Murugaboopathy				

Receiving Date	-	**Name of Faculty Mentor**	P.K.Singh.	**Sign**	P.K.Singh.

➢ Fifth FPR

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Karishma Parashar		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	FPR 5th	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work				■	
Learning capacity/Knowledge up gradation					■
Performance/Quality of work					■
Behaviour/Discipline/Team work				■	
Sincerity/Hard work				■	
Comment on nature of work done/Area/Topic	Karishma going through training modules of Basics of Data Science. She is completing courses on time & performs well.				
OVERALL GRADE (Any one)	**VERY GOOD**				
Name of Industry Mentor	Padmajothi Murugaboopathy				
Signature of Industry Mentor	Padmajothi Murugaboopathy				

Receiving Date		**Name of Faculty Mentor**	P.K.Singh	**Sign**	P.K.Singh

➢ Sixth FPR

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Karishma Parashar		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	FPR 6th	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work					
Learning capacity/Knowledge up gradation					
Performance/Quality of work					
Behaviour/Discipline/Team work					
Sincerity/Hard work					
Comment on nature of work done/Area/Topic	Karishma Now working on Case Study She is completing courses on time & performs well.				
OVERALL GRADE (Any one)	**VERY GOOD**				
Name of Industry Mentor	Padmajothi Murugaboopathy				
Signature of Industry Mentor	Padmajothi Murugaboopathy				

Receiving Date		**Name of Faculty Mentor**	R.K Singh	**Sign**	R.K Singh

➢ Seventh FPR

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Karishma Parashar		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	FPR 7th	
Criterion	**Poor**	**Average**	**Good**	**Very Good**	**Excellent**
Punctuality/Timely completion of assigned work					
Learning capacity/Knowledge up gradation					
Performance/Quality of work					
Behaviour/Discipline/Team work					
Sincerity/Hard work					
Comment on nature of work done/Area/Topic	Karishma working on Case Study Parallely going through module of classification. She is completing courses on time & performs well.				
OVERALL GRADE (Any one)	**VERY GOOD**				
Name of Industry Mentor	Padmajothi Murugaboopathy				
Signature of Industry Mentor	Padmajothi Murugaboopathy				

Receiving Date		**Name of Faculty Mentor**	P.K Singh.	**Sign**	P.K Singh