

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:
Utkarsh Sharma
0901CS181115**

**Faculty Mentor:
Prof.Amit Kumar Manjhvar**

**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 fulfillment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by:

Utkarsh Sharma

0901CS181115

Internship Faculty Mentor:

Prof.Amit Kumar Manjhvar

Submitted to:

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE

GWALIOR - 474005 (MP) est. 1957

MAY-JUNE 2022

Internship Letter:-



Utkarsh Sharma
Jan 18, 2022

Dear Utkarsh,

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
2. Place of Work
Your internship will be administered remotely.
3. Stipend
You will be paid a stipend of INR 30000 (pre-tax) per month during your internship. This will be deposited into your bank account.
4. Benefits
Benefits available to full-time employees such as Provident Fund and Medical Insurance are not applicable to Interns.
5. Leave Entitlements
During your internship period, you are entitled to leave as approved by your manager. Leave cannot be encashed.
6. Safety
The Company is committed to providing a safe working environment for all employees and therefore required to abide by all safety rules and procedures operating within the Company.
7. Conduct
You will be expected to dress appropriately for a business setting. Business casual attire as outlined below is considered appropriate:
 - a. A collared shirt, pants, and shoes for men
 - b. Equivalent Indian or Western business casuals for womenBehaviour unbecoming of a business environment (as perceived the Company) will not be tolerated and might lead to termination of employment.

TIGER ANALYTICS INDIA CONSULTING PVT LTD
(CIN: U74999TN2021FTC146673)
RMZ Millenia Business Park-2, Campus 5(2nd Floor) No. 143, Kandanchavadi,
MGR Road, Perungudi, Chennai 600096. www.tigeranalytics.com

Note:- Company refused to provide Certificate or bonafide before completion of internship therefore, I attached offer letter.



8. Confidentiality

During your employment with the Company, you will make use of Confidential Information in carrying out your duties. Without limitation, "Confidential Information" includes:

1. Information relating to the goods and services and proprietary techniques provided by the Company and clients of the Company
2. All information concerning the business, its methods of operation, marketing and other activities
3. All databases, lists compiled by the company, client proposals, reports, software, algorithms, and computer programs
4. Competitive and financial information concerning the business, which is not in the public domain
5. Information concerning the business of the Company's clients

You must not, whether during employment or after termination of your employment with the Company, without written authority, divulge 'Confidential Information' to anyone other than an employee authorized to receive the information, or use such information for your own personal gain.

9. Inventions and Copyright

You assign to the Company your entire right, title and interest in and to any copyright and any industrial or intellectual property rights in any and all works, designs, computer programs, inventions, processes, concepts, strategies, plans and lists (Confidential Property) which (either solely or jointly with others) you have developed or may develop during and/or as a result of your employment with the Company.

You also agree promptly to disclose to the Company or to its attorneys any and all such Confidential Property developed by you and agree to execute upon demand, at the expense of the Company, all documents which may be desirable to secure to the Company the best copyright, patent or other protection in India and elsewhere and/or rights relating to such Confidential Property.

10. Following End of Internship

a. Confidentiality

You agree that upon termination of your internship with the Company you shall return to the Company:

- a. All documents and any other materials constituting or containing Confidential Property or Confidential Information including, without limitation, customers or contacts, correspondence and other written material relating to Confidential Information or Confidential Property and that you will not retain any such documents or material or copies of such documents or material
- b. Company mobile phone or other electronic telecommunications devices that the company has issued to you. The telephone number of the company owned telecommunications devices will remain property of the company

TIGER ANALYTICS INDIA CONSULTING PVT LTD

(CIN: U74999TN2021FTC146673)

RMZ Millenia Business Park-2, Campus 5(2nd Floor) No. 143, Kandanchavadi,
MGR Road, Perungudi, Chennai 600096. www.tigeranalytics.com



- c. And other property of the Company including, without limitation, security access cards, credit cards, computers and computer software (which must be deleted immediately from any storage device owned by you).

- b. Non-rendering of services to clients

You also agree that you will not, during your employment with the Company or Group and for a period of twelve months after termination, contact, or assist anyone else to contact, any Client of the Company with a view to obtaining business from those Clients; or to persuade those Clients to cease to do business with the Company or to reduce the amount of business which any such Clients has customarily done or is reasonable expected to do with the Company.

The definition of Client includes any prospective client to whom you or the Company has made a formal presentation at any time during the twelve months immediately preceding termination.

You also agree that, if such clients approach you in that period, you will inform them of this agreement and refer them to the appropriate person at the Company.

- c. Non-hire of staff

You also agree that, for six months following termination of your employment with the Company, you will not hire or encourage a third party to hire any member of staff of the Company.

11. Invalidity

In any terms of provisions in this agreement shall be held illegal or unenforceable, in whole or in part, under any enactment or rule of law, such term or provision or part shall to that extent be deemed not to form part of this agreement but the enforceability of the remainder of this agreement shall not be affected.

12. Variation

The terms of this contract of employment may be varied by the Company from time to time. You will be notified of any variations.

13. Adherence to Company Policies

When you join the Company, it will also be a condition of employment that you review and adhere to company policies which you will be notified of subsequently. You agree to adhere to the Company's project financing contracts (e.g. BOT) with the clients.

14. Governing Laws and Jurisdiction

This contract will be governed by the law in force in Chennai, India.

TIGER ANALYTICS INDIA CONSULTING PVT LTD
(CIN: U74999TN2021FTC146673)
RMZ Millenia Business Park-2, Campus 5 (2nd Floor) No. 143, Kandanchavadi,
MGR Road, Perungudi, Chennai 600096. www.tigeranalytics.com



15. Personal Information and Consent

By accepting this offer, you are giving your implicit consent to Tiger Analytics to collect and use your personal information for business purposes. Your personal information may be shared with the Clients and prospective Clients of Tiger Analytics as a part of selection or onboarding process to work in projects. Tiger Analytics will also share your personal information with a third party for carrying out the background verification as required. Tiger Analytics will store your employment, financial and personal information during the period of employment and for Data Retention Period after your separation, as per the data retention policy to comply with statutory requirements.

16. Acceptance

Please sign this letter signifying your acceptance of the appointment and the conditions of service specified in this letter.

We are pleased to welcome you to the Company. If the preceding terms and conditions of your employment with the Company are acceptable to you, please indicate your acceptance by initialing each page and signing the last page of the attached copy and returning it to me.

Regards

G. Pradeep Kumar

Pradeep Gulipalli
General Manager
Tiger Analytics India LLP

TIGER ANALYTICS INDIA CONSULTING PVT LTD
(CIN: U74999TN2021FTC146673)
RMZ Millenia Business Park-2, Campus 5 (2nd Floor) No. 143, Kandanchavadi,
MGR Road, Perungudi, Chennai 600096. www.tigeranalytics.com

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

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

CERTIFICATE


This is certified that **Utkarsh Sharma** (0901CS181115) has submitted the Internship report titled **Data Science Intern** of the work he has done under the mentorship of **Prof. Amit Kumar Manjhvar**, in partial fulfillment 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.


Prof. Amit Kumar Manjhvar

Faculty Mentor

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 fulfillment 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 **Prof.Amit Kumar Manjhvar**, 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.



Utkarsh Sharma

0901CS181115

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 **Prof. Amit Kumar Manjhvar**, 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.



Utkarsh Sharma

0901CS181115

IV Year,

Computer Science and Engineering

ABSTRACT

About Company

Tiger Analytics is pioneering the use of AI and analytics to tackle some of the world's most difficult problems. For several Fortune 500 organizations, we create custom solutions based on data and technology. We have offices in numerous cities across the United States, the United Kingdom, India, and Singapore, as well as a large worldwide virtual workforce. We've won numerous honors, including being named a Forrester Research Leader and being named one of Inc. and Financial Times' fastest-growing IT businesses. We routinely appear on lists of the "Best Analytics Firms." .

About CEO

Tiger Analytics' Founder and CEO is Mahesh Kumar. He founded Tiger Analytics with the goal of using his management science background to help firms achieve better performance using advanced analytics. Mahesh taught at the Smith School of Business and Rutgers Business School before creating Tiger Analytics.

He has undertaken research in the fields of data mining and statistical modelling, and has effectively applied his findings to challenges such as forecasting, pricing, promotions, and consumer segmentation for a variety of firms in a variety of industries.

TABLE OF CONTENTS

TITLE	PAGE NO.
Internship Certificate from Industry	(iii)
Institute Internship Certificate	(viii)
Declaration	(ix)
Acknowledgement	(x)
Abstract	(xi)
Chapter 1: Introduction	1
1.1 Modules	1
1.1.1 Python Programming	1
1.1.2 MS Excel	2
1.1.3 SQL	2
1.1.4 Statistical Concepts	3
1.1.5 Regression Analysis	3
1.1.6 Git and GitHub	3
Chapter 2: Requirement Analysis	4
2.1 Phase 1	4
2.2 Phase 2	4
2.3 Phase 3	4
2.4 Phase 4	4
Chapter 3: System Requirement Specifications	5
Chapter 4: Technology	6
Chapter 5: Project (Work done during internship(Screenshots))	7-10
Chapter 6: FPRs	11-18
References	19

Chapter 1: INTRODUCTION

At the moment, the world is experiencing an IT boom, which means a lot of data is flooding the systems. A lot of this data is useless and useless, but a lot of it is valuable, contains valuable information, and may be utilized to make excellent judgments. To have a greater understanding of reality, this data must be processed. These are the foundations of modern enterprises. They produce a great deal of data yet are unable to interpret and draw conclusions from it. Data science is used in this way. Data analysis provides important findings to help businesses reach new heights. Although Data Science is not limited to business challenges, in today's commercial environment, businesses require Data Science's assistance to move the globe ahead. As a result, this Data Science internship is more focused on the business side of things than on other applications.

For the past four months, I've been working as a Data Science intern at Tiger Analytics. This internship started on January 24, 22. This internship is known as the Springboard Training Program, and it provides all interns with several courses as well as the opportunity to work on real-world business challenges.

The following are the modules covered in this programme:

1.1 Modules

1.1.1 Python Programming

Python is a widely used programming language with a wide range of applications. Data Science is one of them, and python is one of the first steps to success in this field. A comprehensive python course was taught on Udemy in order to obtain proficiency in the language and experience with its use in the Data Science sector. The following themes were covered:

- 1: Basics of Python
- 2: Input/ Output
- 3: Data types
- 4: Conditional Statements
- 5: Loops
- 6: Basics of OOP
- 7: Numpy
- 8: Pandas:
- 9: Scikit Learn
- 10: Matplotlib

1.1.2 MS Excel

The importance of MS Excel is generally recognized in the business world. As a result, it plays a role in Data Science as well. In Data Science, Microsoft Excel is utilized for data storage, data management, and data visualisation. Because not everyone is familiar with technical language, MS Excel is used because it is simple to use and has widespread acceptance and support. It is critical for someone to have a fundamental understanding of MS Excel in order to succeed in the corporate world, whether it is an entry-level position or a GenMan one. The following themes were covered:

- 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

1.1.3 SQL

The acronym SQL refers to Structured Query Language. It's a relational database language that lets you extract information from tables using a set of selection, sorting, and calculation criteria, as well as edit, add, and remove entries. Simply explained, the SQL language is used to write queries that perform various data computations. SQL is used in positions such as data analyst, data engineer, database manager, and so on, and it is a necessary ability to succeed as a data analyst. The following lessons were taught:

- 1: What is Database
- 2: Basics of DBMS
- 3: MySQL workbench
- 4: Database design
- 5: Normalization
- 6: SQL basic commands
- 7: SQL advanced commands
- 8: Functions
- 9: Merging operations

1.1.4 Statistical Concepts

The foundation of data science is statistics. Inferential statistics come into play since data science works with data management, manipulation, and conclusions. Statistics here refers to data calculations employing various testing techniques, validations, modifications, conclusions, and a variety of other topics. Topics covered included:

- 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

1.1.5 Regression Analysis

One of the cornerstones of data science is regression. Through different methodologies, regression is used to analyse, plan, plot, and evaluate numerous data points. The following lessons were taught:

- 1: Introduction to Regression
- 2: Types of regression
- 3: Random Forest Regression
- 4: Models
- 5: Evaluation of model performance

1.1.6 GIT and GitHub

Anyone in the IT industry understands the function and value of git in the life of a developer. Because it was deemed necessary, it was taught alongside github and linux in order to get greater support at various levels of employment.

Chapter 2 : Requirement Analysis

It is critical to examine project demands as they are gathered and throughout the project's life cycle in order for it to be effective. Needs analysis keeps needs in sync with business requirements. A solid needs analysis approach will result in a software package that meets the established business goals. The process of establishing what users anticipate from an application that is being created or upgraded is known as requirement analysis. All efforts conducted to identify the needs of various stakeholders are included in the needs analysis. Analyzing, developing, validating, and maintaining software or system requirements is what needs analysis entails. High-quality requirements are recorded, quantifiable, scalable, and traceable, assisting in the identification of business possibilities, and are described 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.

2.1 Phase 1

Data Cleaning: The whole data has to go through cleaning process, that means removal of junk values, frivolous entries and redundancies. N entries will be checked as per standards and all the issues will be cleared off in order to proceed.

2.2 Phase 2

Data preprocessing: All of the data will be pre-processed in order to make it easy to compute, perform operation upon and calculations. It'll be gone through the standardization to get insights about modelling and gain ideas about approach to solve the questions.

2.3 Phase 3

Data Modelling: Now various questions are dealt upon by formulating models based on samples and rest population data. These models are then used to get the desired results. Modelling requires implementation of various regression algorithms.

2.4 Phase 4

Statistical Analysis: If statistical analysis is required, it is done in the end to mark finishing of the case.

Chapter 3 : System Requirement Specifications

Following are the system requirements for a data analyst:

Hardware Specifications:

Resource	Requirements
Operating systems (64-bit)	Microsoft Windows <ul style="list-style-type: none">• 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 <ul style="list-style-type: none">• 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 <p>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.</p>

Software Requirements:

Resource	Requirements
Port	For more information about ports, see Communication ports and protocols .
Web browsers	<ul style="list-style-type: none">• Mozilla Firefox (latest)• Google Chrome (latest)• Safari on Mac OS -10.x,11• Internet Explorer 11.x
Screen resolution (to view the Console)	1280*1024
Java (for running CLIs)	<ul style="list-style-type: none">• Oracle JRE 1.8.0, build 152 (bundled with IT Data Analytics 11.3.01)• Azul JRE 1.8.0, build 202 <p>Azul 8 Java is supported on version 11.3.02 and later of the product.</p>

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.

Basics:

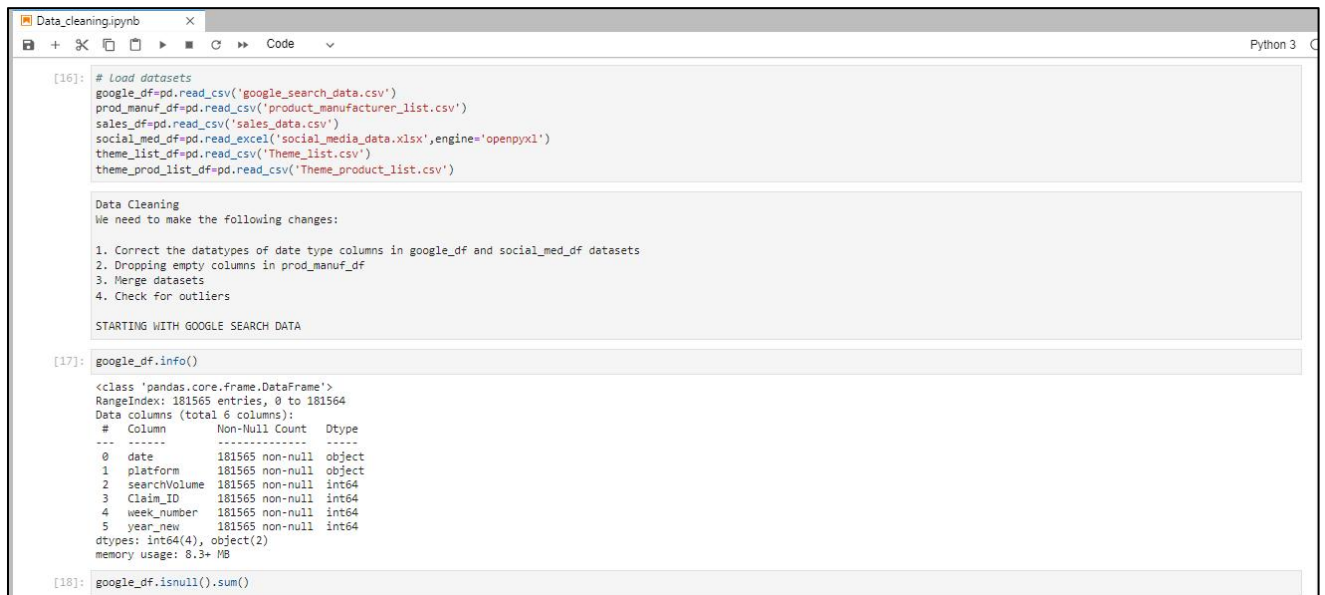
- 1: Python Programming
- 2: SQL
- 3: MS Excel
- 4: Git
- 5: JupyterLab
- 6: Google Collaborator

Intermediate:

- 1: Regression Analysis
- 2: Statistical Analysis

Chapter 5 : Work done during internship (Screenshots)

5.1 Screenshot 1



The screenshot shows a Jupyter Notebook titled 'Data_cleaning.ipynb'. The code in cell [16] loads several datasets: 'google_search_data.csv', 'product_manufacturer_list.csv', 'sales_data.csv', 'social_media_data.xlsx', 'Theme_list.csv', and 'Theme_product_list.csv'. Below the code, a text block titled 'Data Cleaning' lists the tasks: 1. Correct the datatypes of date type columns in google_df and social_med_df datasets, 2. Dropping empty columns in prod_manuf_df, 3. Merge datasets, and 4. Check for outliers. The text 'STARTING WITH GOOGLE SEARCH DATA' is also present. Cell [17] displays the output of 'google_df.info()', showing a DataFrame with 181565 entries and 6 columns: date, platform, searchVolume, Claim_ID, week_number, and year_new. Cell [18] shows the command 'google_df.isnull().sum()'.

```
[16]: # Load datasets
google_df=pd.read_csv('google_search_data.csv')
prod_manuf_df=pd.read_csv('product_manufacturer_list.csv')
sales_df=pd.read_csv('sales_data.csv')
social_med_df=pd.read_excel('social_media_data.xlsx',engine='openpyxl')
theme_list_df=pd.read_csv('Theme_list.csv')
theme_prod_list_df=pd.read_csv('Theme_product_list.csv')

Data Cleaning
We need to make the following changes:

1. Correct the datatypes of date type columns in google_df and social_med_df datasets
2. Dropping empty columns in prod_manuf_df
3. Merge datasets
4. Check for outliers

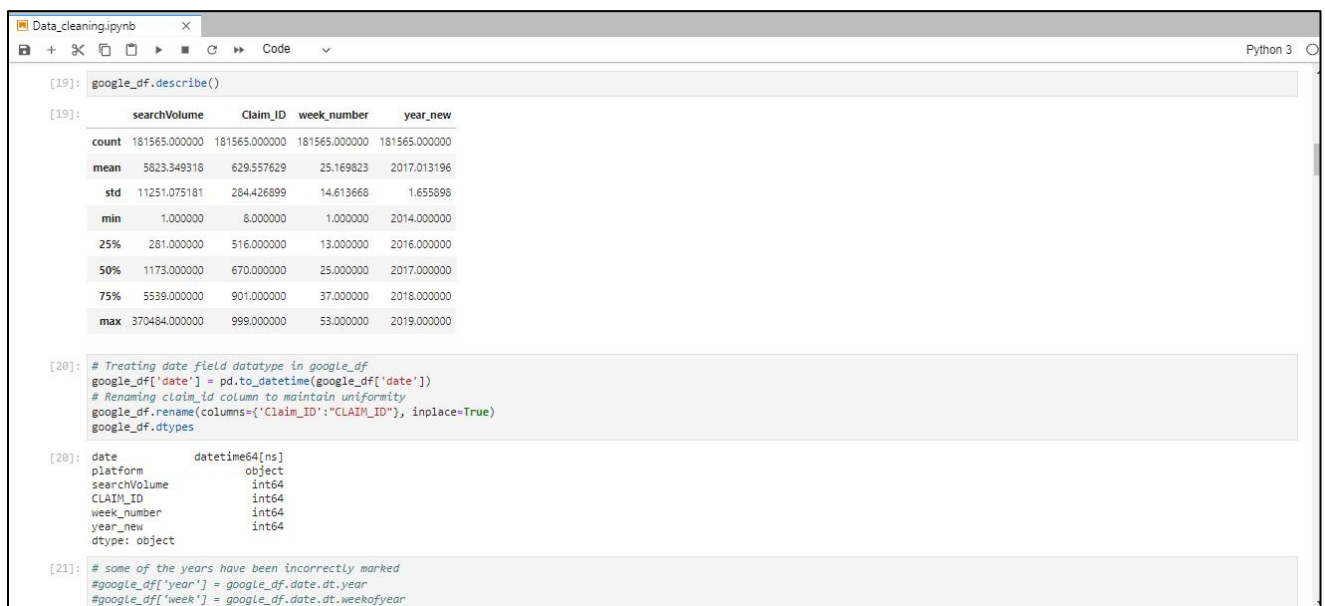
STARTING WITH GOOGLE SEARCH DATA

[17]: google_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181565 entries, 0 to 181564
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype  
---  --
0    date        181565 non-null  object  
1    platform    181565 non-null  object  
2    searchVolume 181565 non-null  int64   
3    Claim_ID    181565 non-null  int64   
4    week_number  181565 non-null  int64   
5    year_new    181565 non-null  int64   
dtypes: int64(4), object(2)
memory usage: 8.3+ MB

[18]: google_df.isnull().sum()
```

5.2 Screenshot 2



The screenshot shows the continuation of the Jupyter Notebook. Cell [19] displays the output of 'google_df.describe()', showing summary statistics for the columns searchVolume, Claim_ID, week_number, and year_new. Cell [20] shows the code for treating the date field datatype and renaming the claim_id column to CLAIM_ID. Cell [21] shows the code for correcting the year and week fields.

```
[19]: google_df.describe()

[19]:      searchVolume  Claim_ID  week_number  year_new
count  181565.000000  181565.000000  181565.000000  181565.000000
mean    5823.349318    629.557629    25.169823    2017.013196
std    11251.075181    284.426899    14.613668     1.655898
min       1.000000       8.000000     1.000000    2014.000000
25%     281.000000     516.000000    13.000000    2016.000000
50%    1173.000000     670.000000    25.000000    2017.000000
75%    5539.000000     901.000000    37.000000    2018.000000
max   370484.000000    999.000000    53.000000    2019.000000

[20]: # Treating date field datatype in google_df
google_df['date'] = pd.to_datetime(google_df['date'])
# Renaming claim_id column to maintain uniformity
google_df.rename(columns={'Claim_ID':'CLAIM_ID'}, inplace=True)
google_df.dtypes

[20]: date          datetime64[ns]
platform         object
searchVolume     int64
CLAIM_ID         int64
week_number      int64
year_new         int64
dtype: object

[21]: # some of the years have been incorrectly marked
#google_df['year'] = google_df.date.dt.year
#google_df['week'] = google_df.date.dt.weekofyear
```


5.3 Screenshot 3

```
2_eda.ipynb
Python 3

Load the cleaned dataset

[ ]: config_path = op.join('conf', 'config.yml')
context = create_context(config_path)
pprint(list_datasets(context))

[ ]: initialize_environment(debug=False, hide_warnings=True)

[ ]: # Load datasets
google_df = load_dataset(context, 'cleaned/google_search')
prod_manuf_df = load_dataset(context, 'cleaned/product_manufact_list')
sales_df = load_dataset(context, 'cleaned/sales')
social_med_df = load_dataset(context, 'cleaned/social_media')
theme_list_df = load_dataset(context, 'cleaned/theme_list')
theme_prod_list_df = load_dataset(context, 'cleaned/theme_prod_list')

EDA and business insights

List of themes available across all data sources

[ ]: google_df = pd.merge(google_df, theme_list_df, on = 'CLAIM_ID', how='inner')

google_list = list(google_df['Claim Name'].unique())
print(google_list)
print("\n\nTotal themes found are : ", len(google_list))

[8]: social_med_df = pd.merge(social_med_df, theme_list_df, on = 'CLAIM_ID', how='inner')

social_media_list = list(social_med_df['Claim Name'].unique())
print(social_media_list)
print("\n\nTotal themes found are : ", len(social_media_list))
```

5.4 Screenshot 4

```
2_eda.ipynb
Python 3

List of themes available across all data sources

[ ]: google_df = pd.merge(google_df, theme_list_df, on = 'CLAIM_ID', how='inner')

google_list = list(google_df['Claim Name'].unique())
print(google_list)
print("\n\nTotal themes found are : ", len(google_list))

[8]: social_med_df = pd.merge(social_med_df, theme_list_df, on = 'CLAIM_ID', how='inner')

social_media_list = list(social_med_df['Claim Name'].unique())
print(social_media_list)
print("\n\nTotal themes found are : ", len(social_media_list))

['tuna', 'allergy', 'convenience - packaging', 'oral health', 'mint', 'dry', 'traditional', 'antioxidant', 'peppermint', 'skin health', 'banana', 'crab', 'mango', 'peanut butter', 'immune health', 'broccoli', 'spinach', 'convenience - ready prepared', 'toffee', 'indulgent & premium', 'organic', 'low fat', 'vitamin/mineral fortified', 'trout', 'low calorie', 'bacon', 'coconut', 'red apple', 'high/source of fiber', 'low sodium', 'no added sugar', 'green bean', 'omega-3', 'bone health', 'green tea', 'cranberry', 'heart health', 'weight management', 'celery', 'strawberry', 'dandelion greens', 'prebiotic', 'ethical - packaging', 'eye health', 'poultry', 'parsley', 'economy', 'cheddar cheese', 'venison', 'ethical - animal/fish & bird', 'low sugar', 'ricotta cheese', 'joint health', 'brain health', 'roquefort cheese', 'no trans fats', 'sugar free', 'diabetic', 'hfcs free', 'low cholesterol', 'kale', 'smoked', 'low gi', 'cashew nut', 'pheasant', 'chili', 'mackerel', 'herring', 'orange', 'lactose free', 'wholegrain', 'added calcium', 'added fiber', 'added iron', 'added protein', 'cod', 'sole', 'sea salt', 'peach', 'butternut', 'fruit', 'nuts', 'anti-aging/aging-well', 'seeds', 'milk chocolate', 'gluten free', 'parmesan cheese', 'carob', 'energy/alertness', 'sports & recovery', 'white cheddar cheese', 'low carb', 'snickerdoodle', 'functional, not specified', 'hemp seeds', 'convenience - consumption', 'kosher', 'soy foods', 'sticks', 'microwavable', 'refill', 'gingerbread', 'single shot', 'seasonal/in-out products', 'gmo free', 'ethical - environment', 'brownie', 'convenience - time saving', 'convenience - easy-to-prepare', 'co-branding', 'ethical - human', 'vegan', 'garlic', 'vegetarian', 'red tomato', 'probiotic', 'ethnic & exotic', 'ethical - not specific', 'halal', 'cloves', 'natural', 'vanilla', 'health (passive)', 'health (active)', 'convenience', 'other', 'rosemary', 'pleasure', 'ethical', 'beans', 'butterscotch', 'choice', 'pepper', 'anchovy', 'herbal', 'fish', 'liver', 'cream', 'beef', 'heart meat', 'quail', 'digestive/gut health', 'cumin', 'pork', 'meat', 'butternut squash', 'high/source of protein', 'herbs', 'crawfish', 'blueberry', 'sweet potato', 'hickory smoked', 'chicken', 'boar', 'dill', 'papaya', 'rabbit', 'duck', 'peanut', 'lamb', 'garden pea', 'berry', 'pumpkin', 'pomegranate', 'french bisque', 'blackberry', 'no additives/preservatives', 'carrot', 'sardine', 'shrimp', 'red raspberry', 'turkey', 'tarragon', 'salmon', 'seafood', 'sausage', 'lobster', 'beef hamburger', 'honey', 'cinnamon', 'cheese', 'yogurt', 'egg', 'vegetables', 'seaweed', 'cherry', 'cola', 'shortbread', 'passion fruit', 'blood orange', 'scallop', 'american gumbo', 'caramel']

Total themes found are : 193

[9]: # Merging on PRODUCT_ID to get all the themes associated with each product
sales_df1 = pd.merge(sales_df, theme_prod_list_df, on = 'PRODUCT_ID', how='inner') # product-theme mapping
sales_df1 = pd.merge(sales_df1, theme_list_df, on = 'CLAIM_ID', how='inner') # Getting claim name
```

5.5 Screenshot 5

```

2_edasipymb
Python 3

Consumer preferences in each of the data sources

• Arranging themes on the basis of the total values of key features in respective datasets.

[10]: # google_df
# google_df['Claim Name'].value_counts().rename_axis('Themes').reset_index(name='Counts').head()

print("The top 5 popular themes in google search data are : ")
google = google_df.groupby('Claim Name')['searchVolume'].sum().sort_values(by='searchVolume', ascending=False)[:15]
print(google[:5])

google = google.reset_index()
google = google['Claim Name'].unique()
google[:5]

The top 5 popular themes in google search data are :
searchVolume
Claim Name
ethical - environment    113502487
shrimp                  77711760
sugar free              74588723
honey                   73957180
health (passive)        56756609

[10]: array(['ethical - environment', 'shrimp', 'sugar free', 'honey',
            'health (passive)'], dtype=object)

[11]: # sales_df
print("Some of the top 5 popular themes in sales data are : ")
sales = sales_df.groupby('Claim Name')['sales_dollars_value'].sum().sort_values(by='sales_dollars_value', ascending=False)[:15]
print(sales[:5])

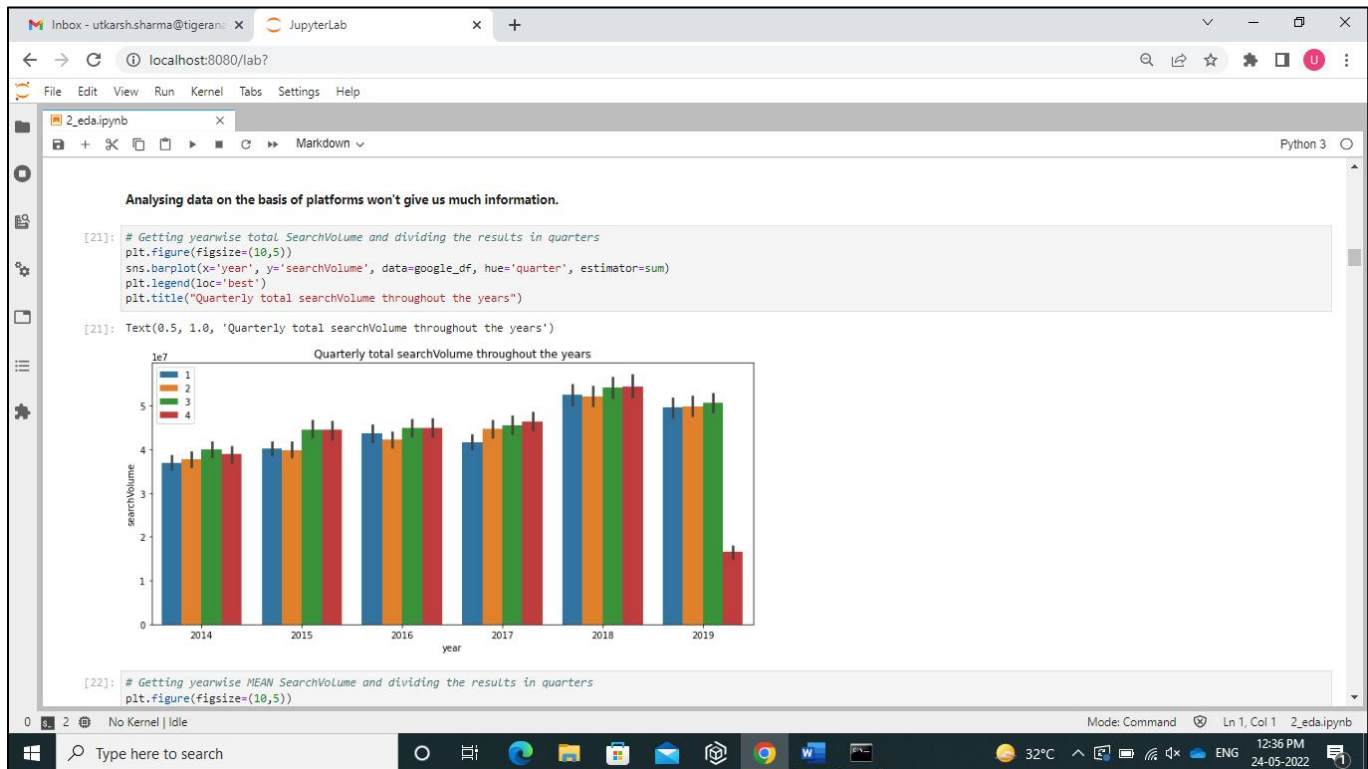
sales = sales.reset_index()
sales = sales['Claim Name'].unique()
sales[:5]

```

5.6 Screenshot 6



5.7 Screenshot 7



Chapter 6: FPRs

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Utkarsh Sharma		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	24/01/22 - 11/02/22	
Criterion	Poor	Average	Good	Very Good	Excellent
Punctuality/Timely completion of assigned work					
Learning capacity/Knowledge up gradation					
Performance/Quality of work					
Behavior/Discipline/Team work					
Sincerity/Hard work					
Comment on nature of work done/Area/Topic	Utkarsh Sharma has started with campus training Program & he is going through the Fundamentals of data science. he is able to complete all the respective assignments on time.				
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	
-----------------------	--	-------------------------------	--	-------------	--

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Utkarsh Sharma		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	12/02/22 - 28/02/22	
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	Utkarsh Sharma has started with campus training Program; he is going through the basics of data science, Stats concepts. he is able to complete all the respective assignments on time.				
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	
-----------------------	--	-------------------------------	--	-------------	--

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Utkarsh Sharma		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	01/03/22 - 15/03/22	
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	Utkarsh is able to complete the training modules & Assignments.				
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	
-----------------------	--	-------------------------------	--	-------------	--

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Utkarsh Sharma		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	15/03/22 - 31/03/22	
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	Utkarsh is able to complete the training modules & Assignments.				
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	
-----------------------	--	-------------------------------	--	-------------	--

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Utkarsh Sharma		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	01/04/22 - 15/04/22	
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	Utkarsh is able to complete the training modules & Assignments.				
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	
-----------------------	--	-------------------------------	--	-------------	--

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Utkarsh Sharma		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	16/04/22 – 30/04/22	
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	Utkarsh is able to complete the training modules & Assignments. He is currently working on case study.				
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	
-----------------------	--	-------------------------------	--	-------------	--

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Utkarsh Sharma		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	01/05/22 – 15/05/22	
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	Utkarsh is able to complete the training modules & Assignments. He is currently working on case study.				
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	
-----------------------	--	-------------------------------	--	-------------	--

FORTNIGHTLY PROGRESS REPORT (FPR) FROM INDUSTRY MENTOR

Name of student	Utkarsh Sharma		Department	Data Science Core	
Industry/Organization	Tiger Analytics		Date/Duration	16/05/22 – 25/05/22	
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	Utkarsh is able to complete the training modules & Assignments. He is currently working on case study.				
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	
-----------------------	--	-------------------------------	--	-------------	--

References:-

- 1.) Tiger Website:- <https://www.tigeranalytics.com/>
- 2.) W3 School:- https://www.w3schools.com/sql/sql_intro.asp
- 3.) Tiger Knowledge Ecosystem:- <https://www.tigeranalytics.com/data-science/>