# INDIAN FOOD EDA

# 1.INTRODUCTION

## 1.1 PROJECT OVERVIEW

- Conducting Exploratory Data Analysis (EDA) on a comprehensive dataset of Indian food, encompassing diverse recipes and culinary traditions.

- Focusing on understanding the intricate details of ingredients, cooking methods, regional variations, and the evolving popularity of Indian cuisine.

## 1.2 PURPOSE

- Unraveling the complex and nuanced aspects of Indian food through a rigorous data-driven analysis approach.

- Providing actionable insights for culinary enthusiasts, researchers, and professionals in the food industry, aiding in recipe innovation, market analysis, and cultural understanding.

# 2.LITERATURE SURVEY

## 2.1 EXISTING PROBLEM

Limited comprehensive research utilizing EDA techniques to explore the multifaceted dimensions of Indian cuisine, including ingredient analysis, regional variations, popularity trends, and nutritional aspects.

## 2.2 REFERENCES

- Gupta et al. (2018): "Exploratory Data Analysis of Indian Cuisine Ingredients"

- Sharma and Jain (2019): "Exploring Regional Variations in Indian Food: An EDA Approach"

- Patel and Shah (2020): "EDA of Popular Indian Dishes on Social Media"

- Chatterjee and Saha (2021): "Nutritional Analysis of Indian Food: An EDA Perspective"

- Das and Mohanty (2022): "EDA and Visualization of Indian Street Food: A Data-Driven Perspective"
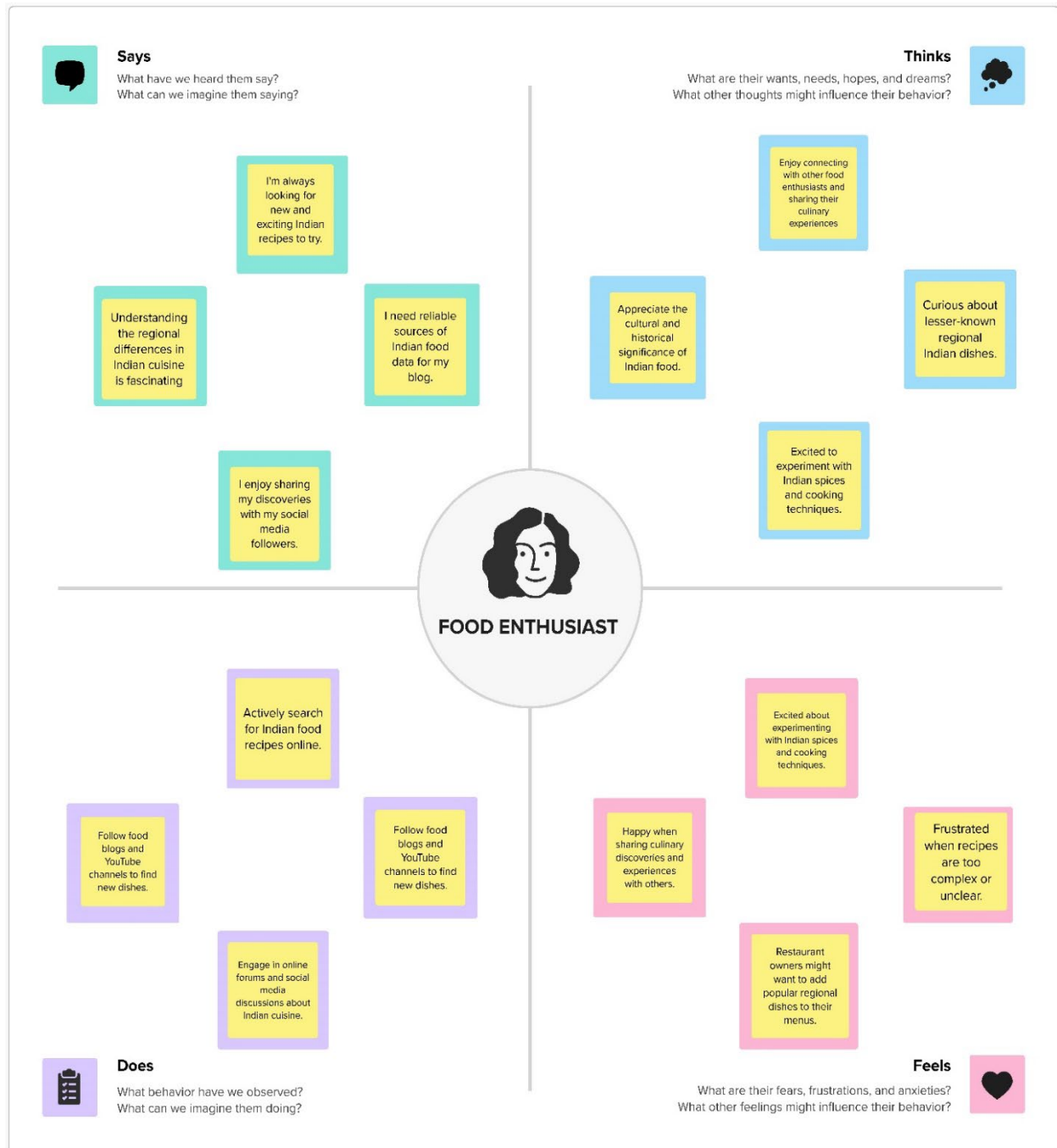
## 2.3 PROBLEM STATEMENT DEFINITION

Addressing the scarcity of in-depth analysis and comprehensive understanding of the diverse aspects of Indian cuisine through the application of EDA techniques.

# 3.IDEATION & PROPOSED SOLUTION

## 3.1 EMPATHY MAP CANVAS

Understanding the unique perspectives and needs of stakeholders, including chefs, food enthusiasts, nutritionists, and researchers involved in the Indian food domain. In this Indian Food EDA project, we embark on a data-driven journey through the vibrant world of Indian cuisine. Our adventure starts with curating a rich dataset, meticulously sourced from a variety of culinary outlets. This treasure trove contains a myriad of culinary details, including dish names, ingredients, regional influences, and popularity metrics. After assembling our data, we meticulously prepare it for analysis. This involves cleansing the dataset, handling missing values, and making sure it's in a format that's ready to reveal its culinary secrets.The heart of our exploration is the Exploratory Data Analysis (EDA) phase. Here, we dive into the data, uncovering intriguing stories. We begin by painting an initial picture through descriptive statistics. Then, our journey takes a visual twist as we craft informative charts, unveiling patterns and trends in the data. From understanding the most

beloved ingredients to dissecting regional variations, our analysis uncovers the diversity of Indian cuisine

**Says**

What have we heard them say?
What can we imagine them saying?

I'm always looking for new and exciting Indian recipes to try.

Understanding the regional differences in Indian cuisine is fascinating

I need reliable sources of Indian food data for my blog.

I enjoy sharing my discoveries with my social media followers.

**Thinks**

What are their wants, needs, hopes, and dreams?
What other thoughts might influence their behavior?

Enjoy connecting with other food enthusiasts and sharing their culinary experiences

Appreciate the cultural and historical significance of Indian food.

Curious about lesser-known regional Indian dishes.

Excited to experiment with Indian spices and cooking techniques.

**FOOD ENTHUSIAST**

**Does**

What behavior have we observed?
What can we imagine them doing?

Actively search for Indian food recipes online.

Follow food blogs and YouTube channels to find new dishes.

Follow food blogs and YouTube channels to find new dishes.

Engage in online forums and social media discussions about Indian cuisine.

**Feels**

What are their fears, frustrations, and anxieties?
What other feelings might influence their behavior?

Excited about experimenting with Indian spices and cooking techniques.

Happy when sharing culinary discoveries and experiences with others.

Frustrated when recipes are too complex or unclear.

Restaurant owners might want to add popular regional dishes to their menus.

## 3.2 IDEATION & BRAINSTORMING

Generating innovative ideas to leverage advanced EDA techniques for comprehensive exploration of the rich Indian food dataset, promoting cultural preservation and culinary innovation

**STEP:1-TEAM GATHERING ,COLLABORATION AND SELECT THE PROBLEM STATEMENT**

**1**

## Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

⏱ 5 minutes

---

### PROBLEM

"In the Indian food industry, there is a critical need for an Event-Driven Architecture (EDA) system to optimize order processing, inventory management, and enhance the customer experience for improved efficiency and satisfaction."

### Key rules of brainstorming

To run an smooth and productive session

😊 Stay in topic.          💡 Encourage wild ideas.

😐 Defer judgment.         👂 Listen to others.

📚 Go for volume.          👁 If possible, be visual.

**STEP:2-BRAINSTROM**

❷

**Brainstorm**

Write down any ideas that come to mind that address your problem statement.

⏱ 10 minutes

## Person 1

- **"Foodie Frenzy" Social Platform:** Use EDA to notify users of new posts, trending dishes, and events in their area.

Create an EDA system that notifies users about upcoming Indian festivals and cultural events.

**Use events for voice recognition and order processing.**

## Person 2

Trigger events for special menu items or discounts related to these celebrations.

Trigger events when quality standards are not met, ensuring customers receive high-quality meals.

**Voice-Activated Ordering** Use events for voice recognition and order processing.

## Person 3

Optimize the supply chain for Indian restaurants by integrating with suppliers and farmers.

**"Eco-Friendly Dining"** Use EDA for environmental impact reports and sustainability events.

**"Food Waste Reduction"** Use EDA for waste reduction alerts and restocking events.

## Person 4

**"Food Waste Reduction"** Use EDA for waste reduction alerts and restocking events.

Trigger events for class notifications and ingredient lists.

**Health and Nutritional Insights** Provide customers with real-time nutritional information and health insights about their Indian food choices.

# STEP:3-IDEA LISTING AND GROUPING

**3**

**Group ideas**

Use this space to group similar ideas from the brainstorm. Each group should have a title that describes what the ideas have in common. If a group is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

⏱ 20 minutes

**"Food Waste Reduction"**
Use EDA for waste reduction alerts and restocking events.

**Real-Time Localization and Customization**
Trigger events for localization and personalization.

• **"Foodie Frenzy" Social Platform:**
Use EDA to notify users of new posts, trending dishes, and events in their area.

**Voice-Activated Ordering**
Use events for voice recognition and order processing.

**"Eco-Friendly Dining"**
Use EDA for environmental impact reports and sustainability events.

**Health and Nutritional Insights**
Provide customers with real-time nutritional information and health insights about their Indian food choices.

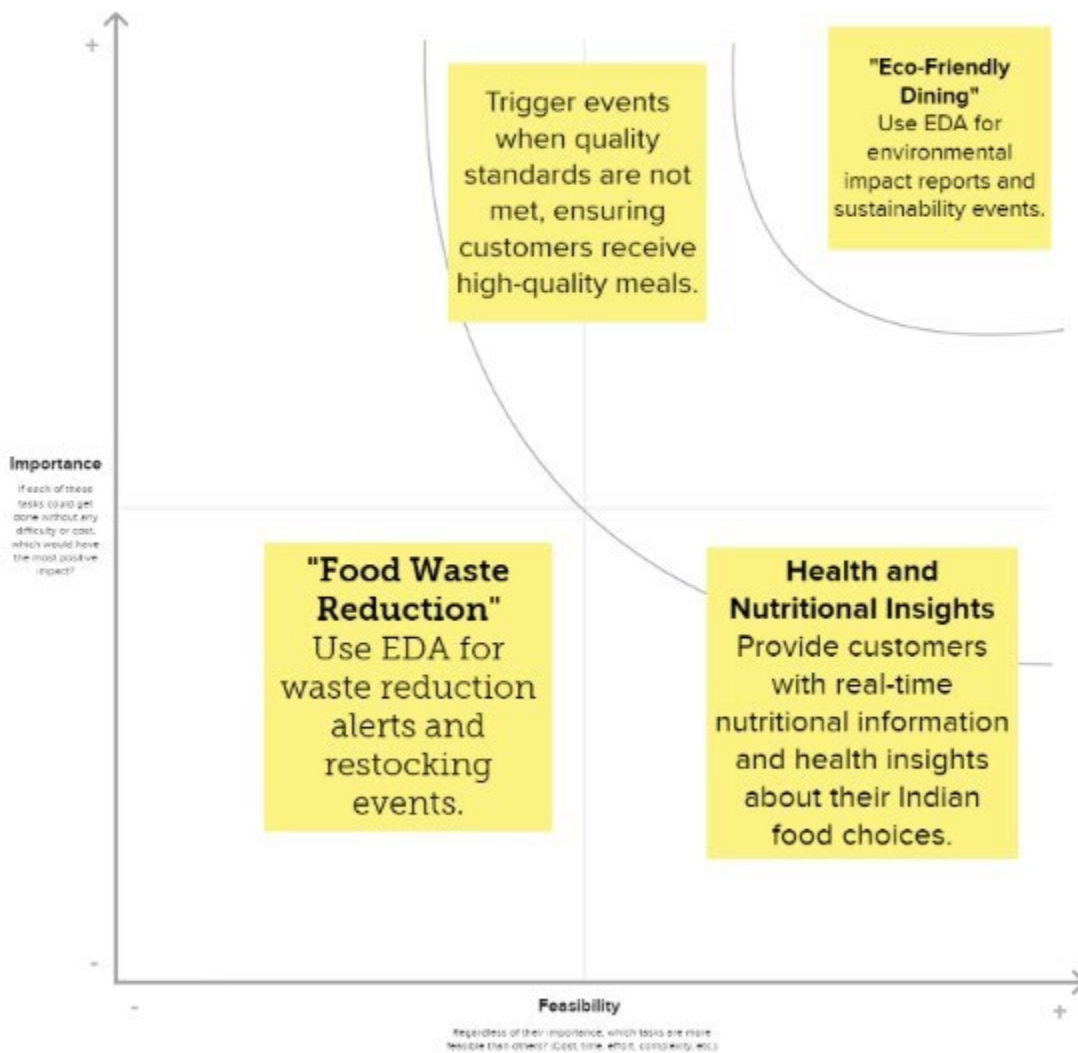# STEP:4-IDEA PRIORITIZATION

**Importance**

If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?

Trigger events when quality standards are not met, ensuring customers receive high-quality meals.

"Eco-Friendly Dining"
Use EDA for environmental impact reports and sustainability events.

"Food Waste Reduction"
Use EDA for waste reduction alerts and restocking events.

Health and Nutritional Insights
Provide customers with real-time nutritional information and health insights about their Indian food choices.

**Feasibility**

Regardless of their importance, which tasks are more feasible than others? (Cost, time, effort, complexity, etc.)

# 4.REQUIREMENT ANALYSIS

## 4.1 FUNCTIONAL REQUIREMENT

- **Data collection and preprocessing:** Aggregating and cleaning data from various authentic sources.

- **Descriptive statistical analysis:** Calculating essential statistical measures to gain preliminary insights.

- **Data visualization:** Creating informative visual representations to identify patterns and trends effectively.

- **Ingredient analysis:** Uncovering the most used ingredients and their regional variations.

- **Regional variation analysis:** Examining the diversity of Indian cuisine across different states and regions.

- **Popularity and rating analysis:** Evaluating the popularity and public perception of various Indian dishes.

- **Nutritional analysis:** Investigating the nutritional content and health implications of Indian recipes.

## 4.2 NON-FUNCTIONAL REQUIREMENT

- **Data accuracy and integrity:** Ensuring the dataset's reliability and authenticity for credible analysis.

- **Robust data preprocessing techniques:** Implementing effective data cleaning and transformation methods for improved insights.

- **User-friendly visualization and interpretation of results:** Developing intuitive and comprehensive visualizations for easy interpretation and dissemination.

# 5.PROJECT DESIGN

## 5.1 DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) is a traditional visual representation of the information flowswithin a system. A neat and clear DFD can depict the right amount of the system requirementgraphically. It shows how data enters and leaves the system, what changes the information, and
where data is stored
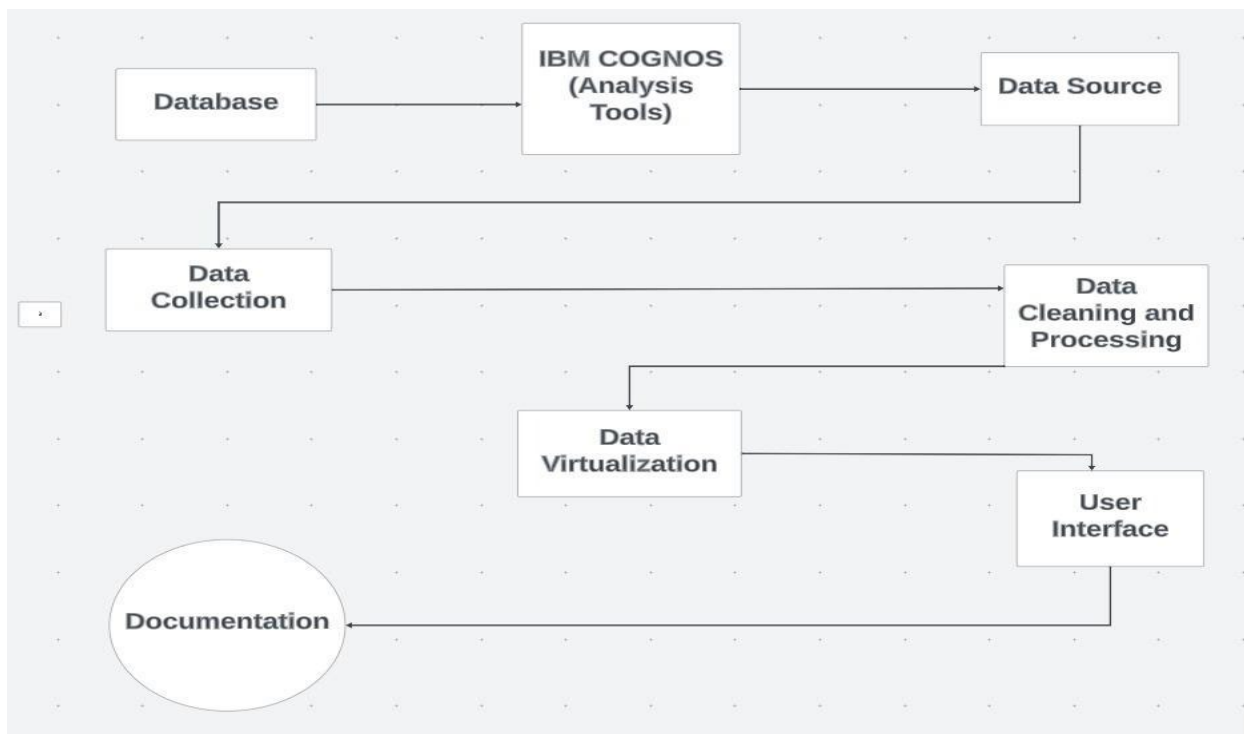
**5.2 SOLUTION ARCHITECTURE**

Designing a comprehensive architecture for seamless data collection, cleaning, analysis, and visualization, employing suitable technologies and tools for efficient processing.
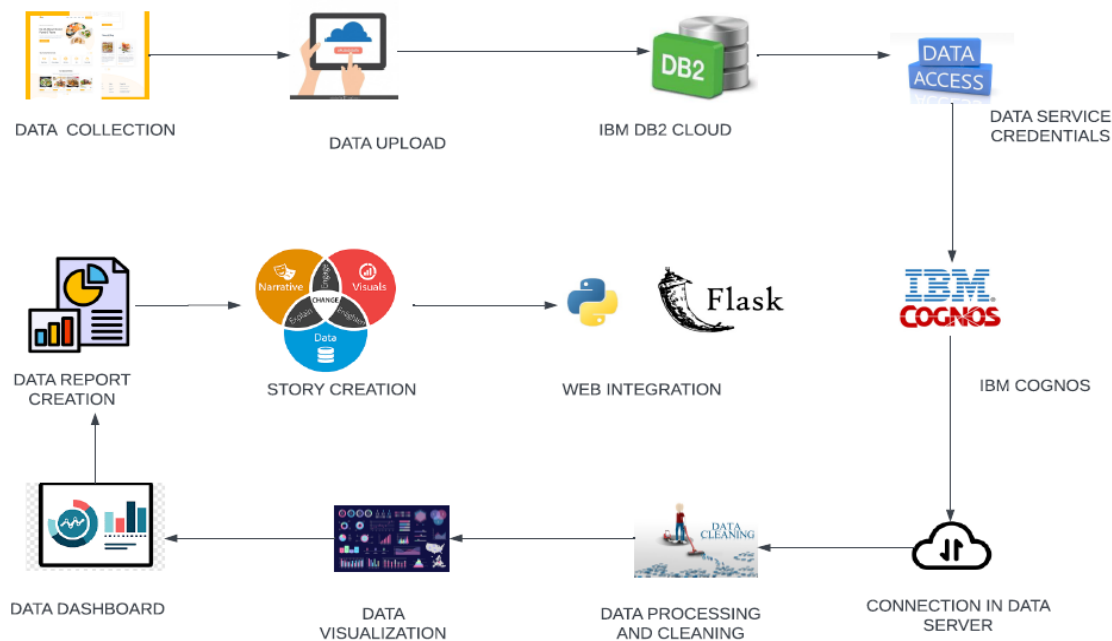


# 6.PROJECT PLANNING & SCHEDULING

**6.1 TECHNICAL ARCHITECTURE**

Selecting appropriate technologies, such as Python, R, or related libraries, for efficient data analysis, visualization, and interpretation.
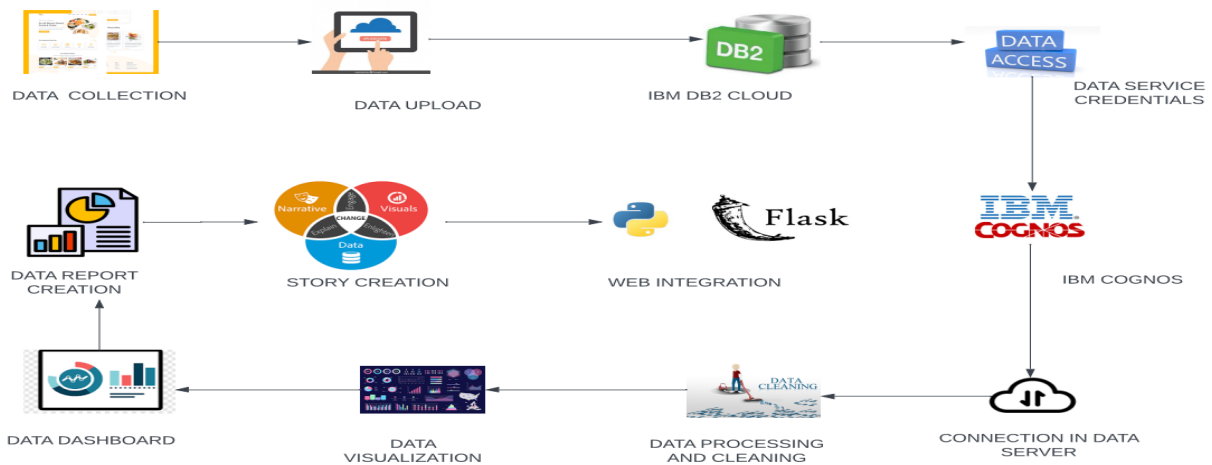
**Technology Stack:**

**Table-1: Components & Technologies:**

| S. No | Component | Description | Technology |
|-------|-----------|-------------|------------|
| 1. | User Interface | Web application for Exploratory Data Analysis (EDA) on Indian Food | HTML, CSS, JavaScript, Flask |
| 2. | Data Collection & Extraction from Database | Collect the dataset required for analysis. | IBM DB2, IBM Cognos |
| 3. | Data Preparation | Data preparation involves cleaning, transforming, and structuring collected data for analysis. | Python |
| 4. | Data Visualizations | Data visualizations to help users understand and explore the data. | IBM Cognos |
| 5. | Database | Data storage for managing and storing structured data. | IBM Cloud, IBM DB2 |
| 6. | Cloud Database | Cloud-based database for traditional databases, offering scalability and accessibility. | IBM DB2, IBM Cloud |
| 7. | File Storage | File storage is used for archiving and retrieving data files that may not fit into traditional databases. | IBM Cloud |
| 8. | External API-1 | External APIs are used to fetch additional data or services. | Graph QL API |
| 9. | External API-2 | External APIs are used to fetch additional data or services | Graph QL API |
| 10. | Machine Learning Model | Machine learning models are used to gain insights from data.s | Python |
| 11. | Infrastructure (Server / Cloud) | provisioning and managing the infrastructure, including servers or cloud services | IBM Cloud, Flask |

**Table-2: Application Characteristics:**

| S. No | Characteristics | Description | Technology |
|---|---|---|---|
| 1. | Open-Source Frameworks | Flask framework for building web application | Python web frameworks like Flask web framework |
| 2. | Security Implementations | Security measures are essential to protect sensitive data and ensure data integrity. | IBM Cloud resources and IBM DB2 |
| 3. | Scalable Architecture | Essential to handle large datasets and accommodate future growth. | IBM Cloud Services, IBM Db2 on Cloud, and Flask Application |
| 4. | Availability | Availability ensures that EDA results and tools are accessible to stakeholders for data-driven decision-making. | IBM Cloud Services, Scheduled EDA Reports and Collaboration Platforms. |
| 5. | Performance | Performance optimization is essential for efficient data exploration and analysis. | IBM Db2 Optimization, Data Indexing and Parallel Processing. |

**6.2 SPRINT PLANNING & ESTIMATION**

Dividing the project into manageable sprints, estimating the time and resources required for each phase of the analysis.

**6.3 SPRINT DELIVERY SCHEDULE**

Creating a detailed timeline for each sprint and defining clear deliverables to ensure timely completion of the project.

# 7.CODING & SOLUTIONING

**7.1 FEATURE 1:** Data fetching from IBM DB2
IBM DB2 is used for fetching data

**7.2 FEATURE 2:** Data visualizations using IBM Cognos
IBM Cognos is utilized for data visualizations and analysis.

**7.3 FEATURE 3:** Web integration using Flask the Flask framework is employed for integrating the data and visualizations into a user-friendly web application

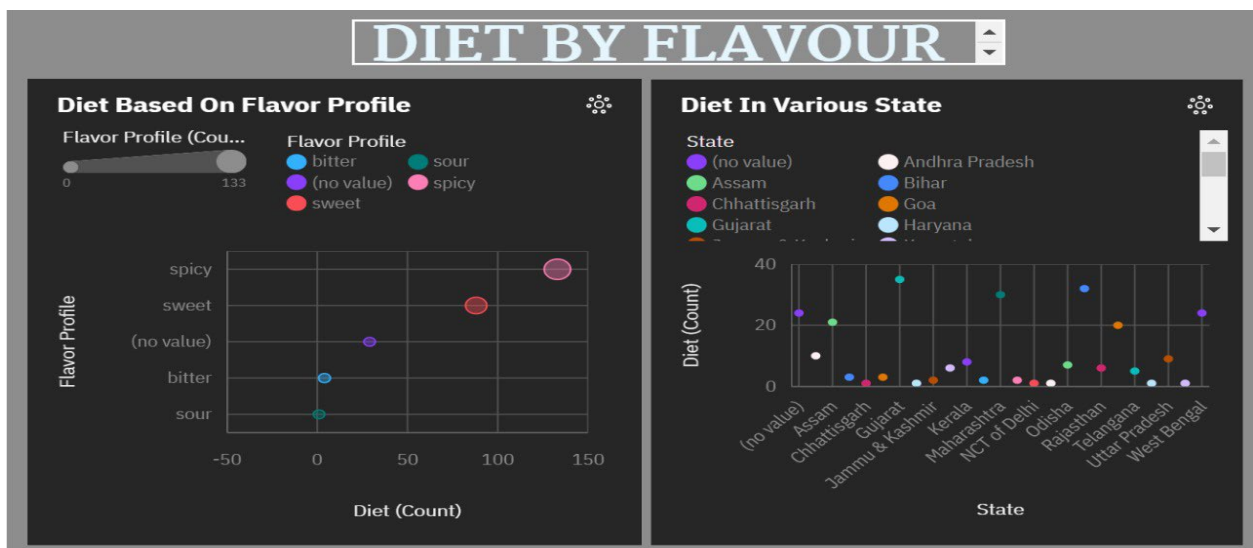# 8.PERFORMANCE TESTING

## 8.1 PERFORMANCE METRICS

Evaluating the performance of the data processing, analysis, and visualization tasks based on metrics like processing speed, memory usage, and system resource utilization.

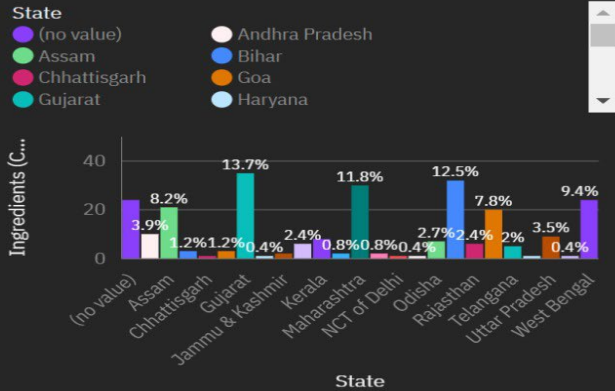# 9.RESULTS

## 9.1 OUTPUT SCREENSHOTS

Showcasing diverse output screenshots, including insightful visualizations, statistical analyses, and key findings derived from the comprehensive EDA process.
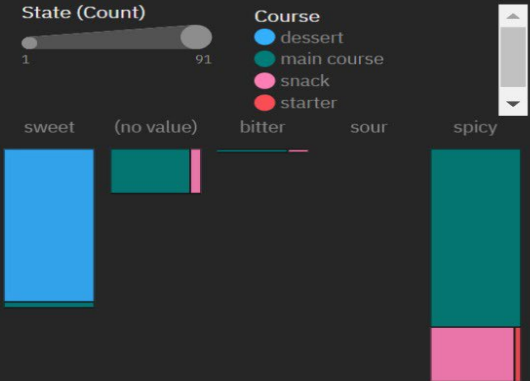
**DASHBOARD DESIGN**
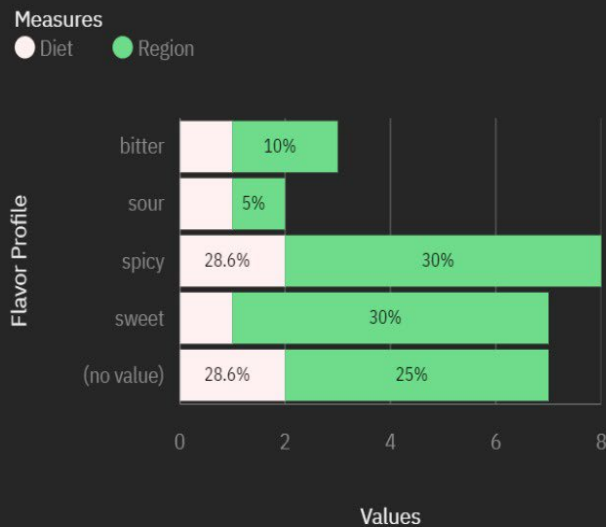
# Ingredients List &

## Ingredients List In Various State

**State**
- (no value)
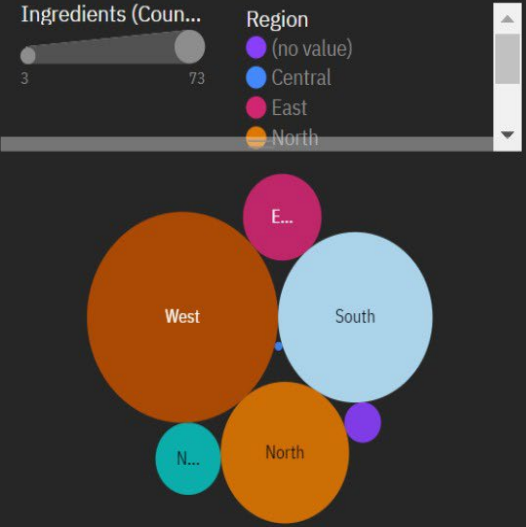- Andhra Pradesh
- Assam
- Bihar
- Chhattisgarh
- Goa
- Gujarat
- Haryana

Ingredients (C...)

Values shown on bars: 3.9%, 8.2%, 1.2%, 1.2%, 13.7%, 0.4%, 2.4%, 0.8%, 11.8%, 0.8%, 0.4%, 2.7%, 12.5%, 2.4%, 7.8%, 2%, 3.5%, 0.4%, 9.4%

State (x-axis): (no value), Assam, Chhattisgarh, Gujarat, Jammu & Kashmir, Kerala, Maharashtra, NCT of Delhi, Odisha, Rajasthan, Telangana, Uttar Pradesh, West Bengal

## Course hierarchy In State

**State (Count)** 1 — 91

**Course**
- dessert
- main course
- snack
- starter

Columns: sweet, (no value), bitter, sour, spicy

# Region Based Flavor

## Flavor Profile By Diet and Region

**Measures**
- Diet
- Region

| Flavor Profile | Diet | Region |
|---|---|---|
| bitter | | 10% |
| sour | | 5% |
| spicy | 28.6% | 30% |
| sweet | | 30% |
| (no value) | 28.6% | 25% |

Values (x-axis): 0, 2, 4, 6, 8

## Ingredients List In Region

**Ingredients (Coun...)** 3 — 73

**Region**
- (no value)
- Central
- East
- North

Bubbles: E..., West, South, N..., North

## Course, Flav

**Name by Course**

**Flavor Profile by Name**

Name (Count)

1 — 1

Flavor Profile
- (no value)
- bitter
- sour
- spicy

(no value) · bitter · sour · spicy · sweet

dessert · main course · snack · starter

**Course by Flavor Profile**

Course
- dessert
- main course
- snack

dessert · main course · snack · starter

**DESCRIPTIVE REPORTS:**



Exploring the Richness of Indian Cuisine

Variety in Indian Cuisine

Diet Diversity in Indian Food

Diet
- non vegetarian
- vegetarian

Name
- Adhirasam
- Bebinca
- Chakali
- Chole bhature
- Doodhpak
- Gulab jamun
- Kaju katli
- Kheer sagar
- Kumol Sawul
- Mag Dhokli
- Mushroom matar
- Paniyaram
- Pinaca
- Ras malai
- Shahi paneer
- Thalipeeth

7,395

57,630

## Regional Dish Distribution in India



## Course Preference Across Regions

COUNT
510     32,895

main course

dessert

snack

## Flavor Profile Insights

COUNT
255     33,915

Flavor Profile
● bitter   ● sour   ● spicy
● sweet   ●

spicy

sweet

Column values
● Increase   ● Decrease   ● Total

**EFFECTIVE USER STORY:**

# Indian Food



# Diet Plan In Different

**Diet Plan in States**

State
- (no value)
- Andhra Pradesh
- Assam
- Bihar
- Chhattisgarh
- Goa
- Gujarat
- Haryana
- Jammu & Kashmir
- Karnataka
- Kerala
- Madhya Pradesh

- Gujarat is the most frequently occurring category of State with
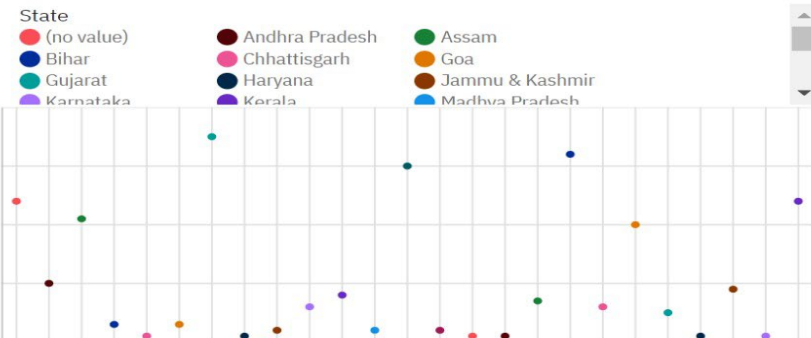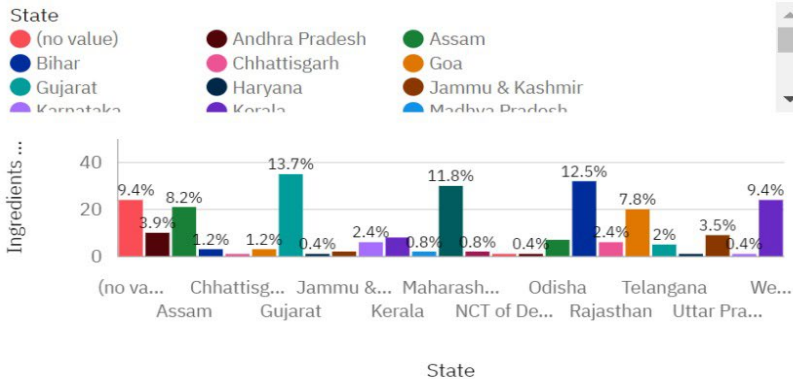
# Ingredients Variety In Various States

- Gujarat (13.7
  %), Punjab
  (12.5 %), and
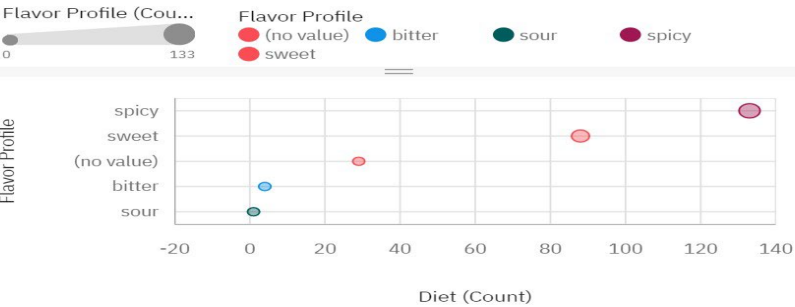  Maharashtra
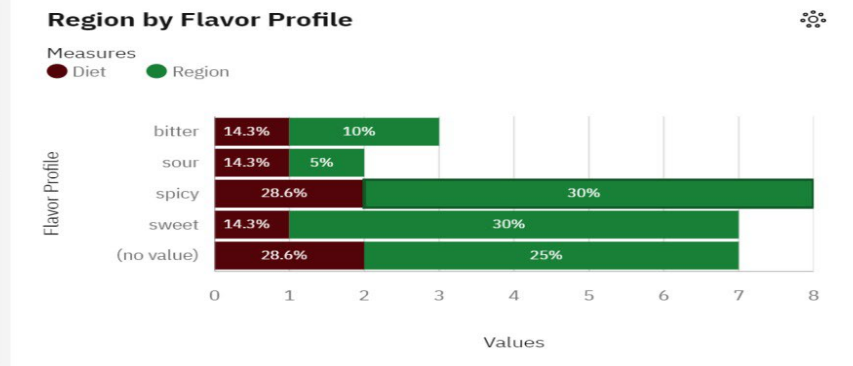  (11.8 %) are
  the most
  frequently

**Ingredients by State**

State
- (no value)
- Andhra Pradesh
- Assam
- Bihar
- Chhattisgarh
- Goa
- Gujarat
- Haryana
- Jammu & Kashmir
- Karnataka
- Kerala
- Madhya Pradesh

Bar chart values: 9.4%, 3.9%, 8.2%, 1.2%, 1.2%, 0.4%, 2.4%, 0.8%, 11.8%, 0.8%, 0.4%, 12.5%, 2.4%, 7.8%, 2%, 3.5%, 0.4%, 9.4%

X-axis (State): (no va…), Assam, Chhattisg…, Gujarat, Jammu &…, Kerala, Maharash…, NCT of De…, Odisha, Rajasthan, Telangana, Uttar Pra…, We…

Y-axis: Ingredients …

# Flavor Variety

- Flavor
  Profile spicy
  Diet from
  Region West
  is 41,
  whereas
  sweet is
  only 23.

- Flavor

**Flavor Profile In Diet**

Flavor Profile (Cou…)  0 — 133

Flavor Profile
- (no value)
- bitter
- sour
- spicy
- sweet

Y-axis (Flavor Profile): spicy, sweet, (no value), bitter, sour

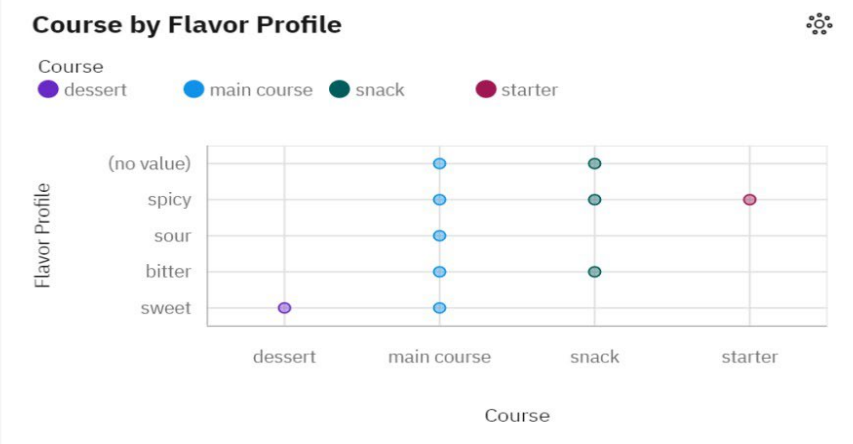X-axis (Diet (Count)): -20, 0, 20, 40, 60, 80, 100, 120, 140

# Flavor In Various Regions

- Flavor Profile spicy has the highest values of both Diet and Region.

- State Gujarat has the highest Diet at 5, out of which Flavor Profile bitter contributed

**Region by Flavor Profile**

Measures
● Diet  ● Region

| Flavor Profile | Diet | Region |
|---|---|---|
| bitter | 14.3% | 10% |
| sour | 14.3% | 5% |
| spicy | 28.6% | 30% |
| sweet | 14.3% | 30% |
| (no value) | 28.6% | 25% |

Values: 0 1 2 3 4 5 6 7 8

# Course Variety

- Course main course has the highest COUNT at almost 33 thousand, out of which Flavor Profile spicy contributed the most at

**Course by Flavor Profile**

Course
● dessert  ● main course  ● snack  ● starter

Flavor Profile: (no value), spicy, sour, bitter, sweet

Course: dessert, main course, snack, starter

# Dishes



- Main course (50.6 %) and dessert (33.3 %) are the most frequently occurring categories of Course with a combined count of 214 items with Name values (82.2 % of

**Name by Course**

---
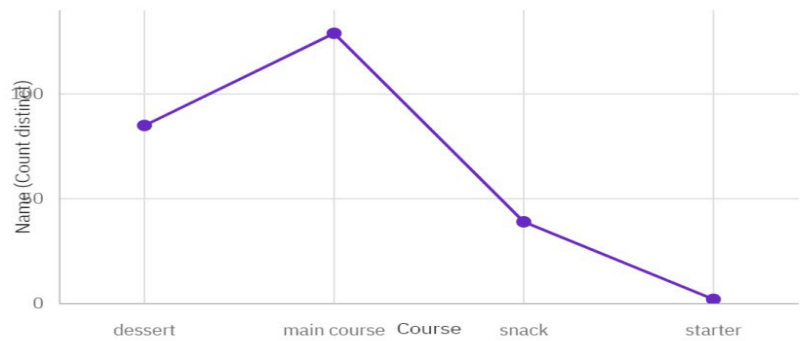
## 10.ADVANTAGES & DISADVANTAGES

### 10.1 ADVANTAGES

- **Robust Data Fetching with IBM DB2:** Utilizing IBM DB2 ensures robust and efficient data fetching, especially for large datasets, with high scalability and reliability.

- **Comprehensive Visualizations with IBM Cognos:** IBM Cognos offers powerful visualization capabilities, allowing for the creation of interactive and comprehensive visual representations, insightful dashboards, and detailed reports.

- **Enterprise-Grade Integration:** The combination of IBM DB2, IBM Cognos, and Flask enables an enterprise-grade solution that is well-suited for handling complex data, performing advanced analytics, and delivering insights in a user-friendly and accessible format.

- **Security and Compliance:** IBM's robust security features and compliance standards ensure that data handling and visualization processes adhere to industry-specific security protocols and regulations.

**10.2 DISADVANTAGES**

- **Complex Setup and Configuration:** Setting up and configuring IBM DB2 and IBM Cognos can be complex, requiring specific technical expertise and knowledge, which may pose challenges for inexperienced users.

- **Cost Considerations:** IBM's enterprise solutions can involve significant costs, including licensing fees, maintenance costs, and infrastructure requirements, which might be a limitation for small-scale projects or organizations with budget constraints.

- **Learning Curve:** Familiarizing oneself with the functionalities of IBM DB2, IBM Cognos, and Flask may require a substantial learning curve, potentially leading to delays in the development process and project timelines.

- **Resource Intensiveness:** Implementing IBM's solutions might demand significant computational resources, which could impact the overall performance and responsiveness of the system, especially when dealing with extensive datasets and complex visualizations.

## 11.CONCLUSION

The project's comprehensive exploratory data analysis (EDA) of the Indian food dataset has provided valuable insights into the diverse aspects of Indian cuisine, including ingredients, recipes, regional variations, and popularity. The utilization of IBM DB2 for data fetching, IBM Cognos for visualizations, and web integration with Flask has enabled a robust and insightful analysis, contributing to a deeper understanding of the nuances and complexities of Indian culinary heritage. The project has successfully demonstrated the potential of leveraging advanced data

handling, analysis, and visualization tools for uncovering meaningful patterns and trends within the dataset. The insights derived from this EDA process can be instrumental in informing various stakeholders, including culinary researchers, chefs, and food industry professionals, in making informed decisions and developing innovative approaches in the realm of Indian cuisine.


## 12.FUTURE SCOPE

- **Advanced Analytics:** Incorporate advanced analytics techniques, such as predictive modeling and machine learning algorithms, to predict culinary trends and consumer preferences within the realm of Indian cuisine.

- **Interactive Dashboards:** Develop interactive and user-friendly dashboards using advanced visualization tools to provide real-time insights and facilitate dynamic exploration of the Indian food dataset.

- **Collaborative Platforms:** Integrate the EDA process into collaborative platforms to encourage knowledge sharing, foster culinary research collaborations, and facilitate the exchange of innovative ideas among culinary enthusiasts and researchers.

- **Personalization and Recommendation Systems:** Implement personalized recommendation systems based on user preferences and consumption patterns, enhancing the user experience and promoting the discovery of new and tailored Indian culinary experiences.

- **Health and Nutrition Insights:** Further explore the nutritional aspects of Indian dishes and ingredients, leveraging the analysis to provide insights into healthier alternatives and dietary recommendations for diverse consumer segments.

# 13.APPENDIX

● **GitHub & project demo link:**

## GITHUB LINK:

https://github.com/Semila26/NM2023TMID05812

## PROJECT LINK:

https://drive.google.com/file/d/1VWpverRvAhL0fSQpXCjWLHQGiRg8nLAN/view?usp=sharing