NAAN MUDHALVAN – DATA ANALYTICS WITH TABLEAU

PROJECT TITLE: INDIAN FOOD EDA

TEAM MEMBERS

1.Joseph James

2.Thunga Naveen

3.Nikhil p s

4.Ranjith

1.INTRODUCTION

The aim of this project is to perform Exploratory Data Analysis (EDA) on a dataset related to Indian food. Exploratory Data Analysis is a crucial step in data analysis that helps in understanding the dataset, discovering patterns, and extracting insights. In this project, we will explore various aspects of Indian cuisine, such as ingredients, recipes, regional variations, and popularity.

1.1 PROJECT Purpose

The project aims to analyse different Indian dishes and infer about the time taken for preparing each dish with help of Exploratory Data Analysis (EDA) on a dataset created based on Indian food varieties, thus pictorially representing the preparation process as slow, medium, and fast.

1.2 PROJECT OVERVIEW

Exploratory Data Analysis (EDA) on Indian Food

DATA COLLECTION

The first step is to collect a comprehensive dataset on Indian food. This dataset can be obtained from various sources such as recipe websites, food blogs, culinary books, or public repositories. The dataset should ideally include information such as the name of the dish, ingredients, cooking methods, region, popularity, and any other relevant attributes.

DATA CLEANING AND PREPROCESSING

Once the dataset is obtained, it needs to be cleaned and pre-processed before analysis. This involves handling missing values, removing duplicates, standardizing formats, and transforming the data into a suitable structure for analysis. For example, ingredient names and measurements might need to be standardized, and textual data may require text cleaning techniques like removing stop words, punctuation, and converting to lowercase.

EXPLORATORY DATA ANALYSIS (EDA)

After cleaning the dataset, the EDA process begins. The analysis can involve various steps and techniques, including:

1.   Descriptive Statistics: Calculating basic statistics such as mean, median, mode, range, and standard deviation of different attributes. This provides an initial overview of the dataset.

2.   Data Visualization: Creating visual representations of the data to identify patterns, trends, and relationships. This can be done using charts, graphs, heatmaps, scatter plots, and geographical maps. For example, visualizing the distribution of recipes across different regions or plotting the popularity of dishes over time.

3.   Ingredient Analysis: Analyzing the most used ingredients, their frequency, and variations across different regions. This can help identify key ingredients that define Indian cuisine and understand regional preferences.

4.   Regional Variations: Investigating the diversity of Indian cuisine across different regions and states. This can include analyzing the prevalence of specific dishes, spices, cooking styles, and regional specialties.

5.   Popularity and Ratings: Analyzing the popularity and ratings of Indian dishes or cuisines. This can involve examining user reviews, ratings, social media mentions, or any available popularity metrics to identify popular dishes or trends.

6.   Nutritional Analysis: Exploring the nutritional content of Indian dishes, such as calories, macronutrients, and common ingredients contributing to specific nutritional values. This can help understand the health aspects of Indian cuisine.

TECHNICAL ARCHITECTURE
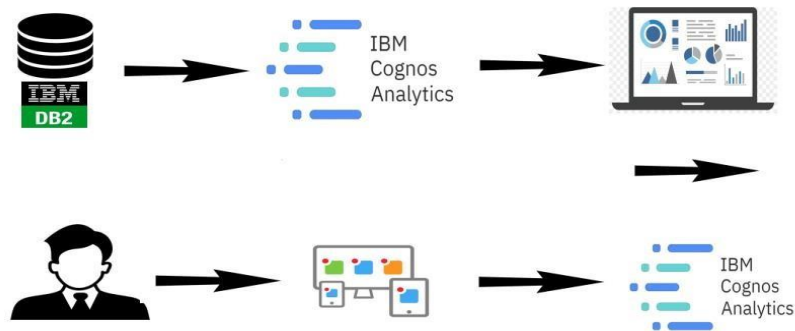
PROJECT FLOW

To accomplish this, we have to complete all the activities listed below, Define Problem / Problem Understanding ○ Specify the business

problem or Business requirements ○ Literature Survey ○

Social or Business Impact.

• **Data Collection & Extraction from Database**

Collect the dataset

Connect IBM DB2 with IBM Cognos

• **Data Preparation** ○ Prepare the Data for Visualization

• **Data Visualizations**
○ No of Unique **Visualizations**

• **Dashboard**

○ Responsive and Design of Dashboard

• **Story**

○ No of Scenes of Story

- **Report**
  - o Creating a report

- **Performance Testing**

  o Amount of Data Rendered to DB ' o Utilization of Data

  Filters o No of Calculation

  Fields o No of Visualizations/ Graphs

- **Web Integration** o Dashboard and Story embed with UI
  With Flask

- **Project Demonstration & Documentation** o Record
  explanation video for project end to end solution   o
  Project Documentation - Step by step project development
  procedure

2.<u>Literature Survey</u>

LITERATURE REVIEW: EXPLORATORY DATA ANALYSIS (EDA) ON INDIAN FOOD

INTRODUCTION

Exploratory Data Analysis (EDA) has become an essential tool in understanding and extracting insights from large datasets. In the context of Indian food, EDA plays a crucial role in unraveling the complexities of this diverse and rich culinary heritage. This literature review explores existing research and studies that have employed EDA techniques to analyze Indian food datasets, shedding light on the ingredients, regional variations, popularity, and nutritional aspects of Indian cuisine.

=

he literature review demonstrates the growing interest in applying EDA techniques to explore Indian food datasets. These studies have provided valuable insights into the ingredients, regional variations, popularity, and nutritional aspects of Indian cuisine. EDA has proven instrumental in unravelling the intricacies of Indian food, shedding light on the diversity, flavours, and cultural significance associated with this culinary heritage. Future research can further leverage EDA to explore new dimensions of Indian food, such as the impact of globalization, dietary preferences, and the fusion of traditional and modern culinary practices.

2.1 Define Problem / Problem Understanding

Activity 1:

Specify the business problem - Refer Project Purpose Activity

2:

Business requirements:

The business requirements for this project would likely include:

- Data collection:

  The first requirement is to collect relevant dataset from Kaggle based on the region, ingredients, cuisines, courses, and diet in order to compare food varieties from different regions of India and the World.

- Data cleaning and preparation:

  The collected data must be cleaned and processed to ensure it is suitable for analysis. This may involve removing irrelevant information, correcting inconsistency, and missing values, and transforming the data into a format that is compatible with the analysis tools.

- Data analysis:

  The data must be analyzed to uncover meaningful insights. This could involve using techniques such as descriptive statistics, regression analysis, and data visualization to gain a deeper understanding of the data.

- Report creation:

  The insights and findings from the data analysis must be presented in a comprehensive report that includes visualizations and data tables. The report must be well organized and easy to understand, with clear and concise explanations of the results.

Activity 4: Social or Business Impact

1. Social Impact:

   Health and Nutrition Awareness: EDA on Indian food datasets facilitates the exploration of nutritional aspects, enabling individuals to make informed dietary choices. By analysing calorie content, macronutrients, and ingredient compositions, EDA promotes health and nutrition awareness. This knowledge empowers individuals to adopt healthier eating habits and make informed decisions about their diet.

2. Business Model/Impact:

   Restaurant Menu Planning: EDA on Indian food datasets helps restaurants and food establishments understand customer preferences and adapt their menus accordingly. By analysing popular dishes, regional variations, and emerging trends, businesses can tailor their offerings to meet consumer demands. This improves customer satisfaction and drives business growth.

EDA on Indian food datasets has a profound social and business impact. It promotes the preservation of culinary heritage, encourages cultural exchange, and raises awareness of health and nutrition. In the business realm, EDA drives menu planning, product development, marketing strategies, and supply chain optimization.
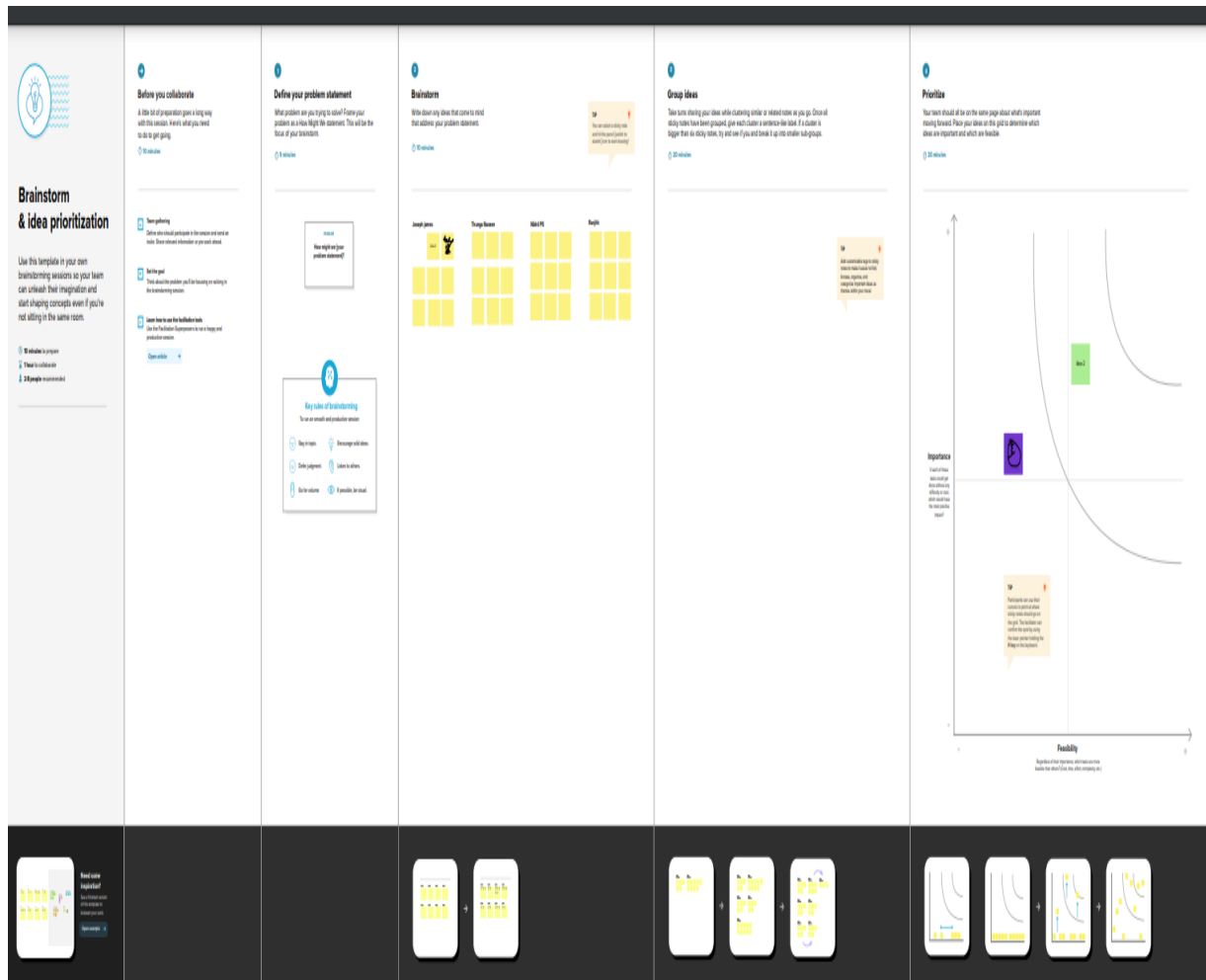
3.IDEATION &PROPOSED SOLUTION

3.1 Empathy Map Canvas
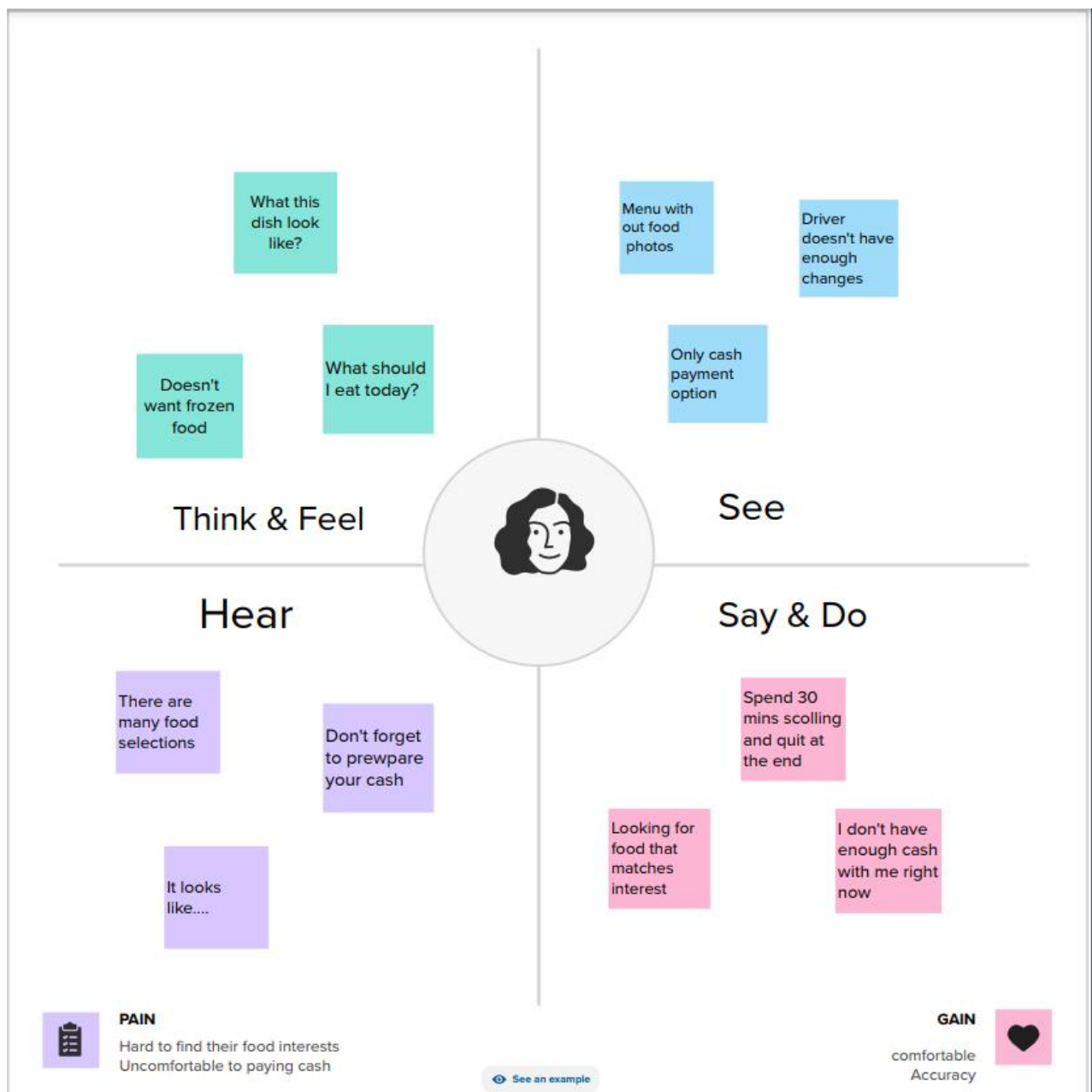
3.2 Ideation & Brainstorming

4. PROJECT DESIGN

4.1 Data Flow Diagram & User Stories

Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

Example: DFD Level 0 (Industry Standard) **Example: (Simplified)**

Flow
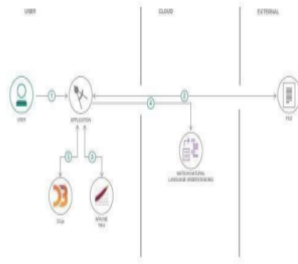
1. User configures credentials for the Watson Natural Language Understanding service and starts the app.
2. User selects data file to process and load.
3. Apache Tika extracts text from the data file.
4. Extracted text is passed to Watson NLU for enrichment.
5. Enriched data is visualized in the UI using the D3.js library.
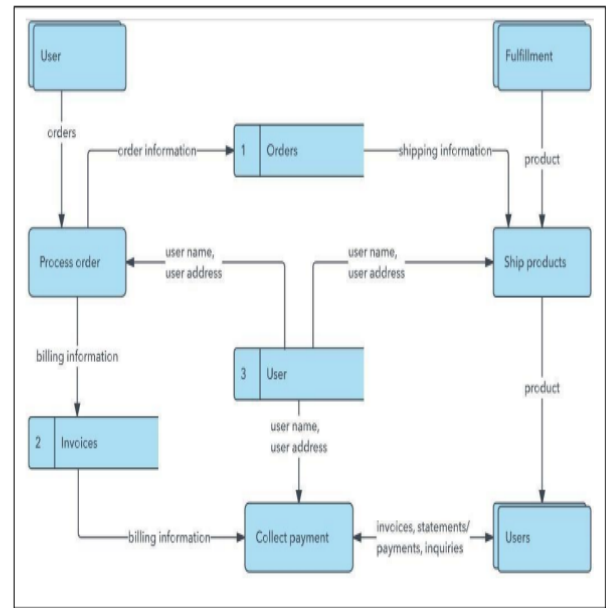
User Stories

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard					
Customer (Web user)						
Customer Care Executive						
Administrator						

4.2 Solution Architecture

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.

- Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.

- Define features, development phases, and solution requirements.

- Provide specifications according to which the solution is defined, managed, and delivered.
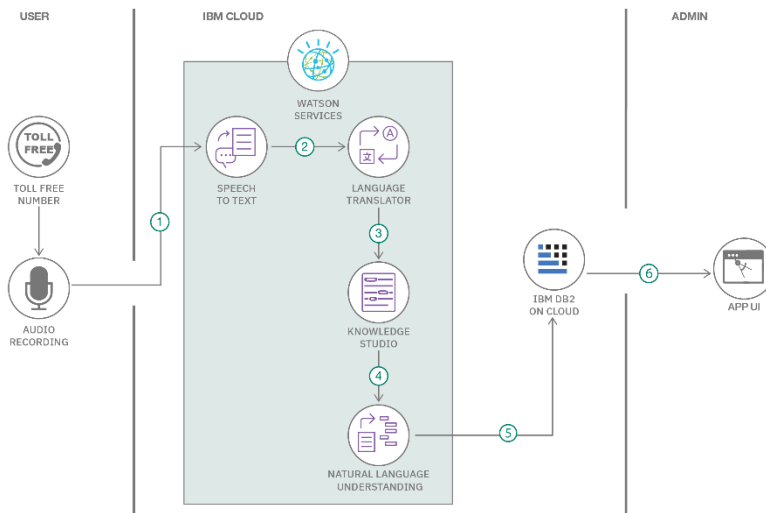
Example - Solution Architecture Diagram
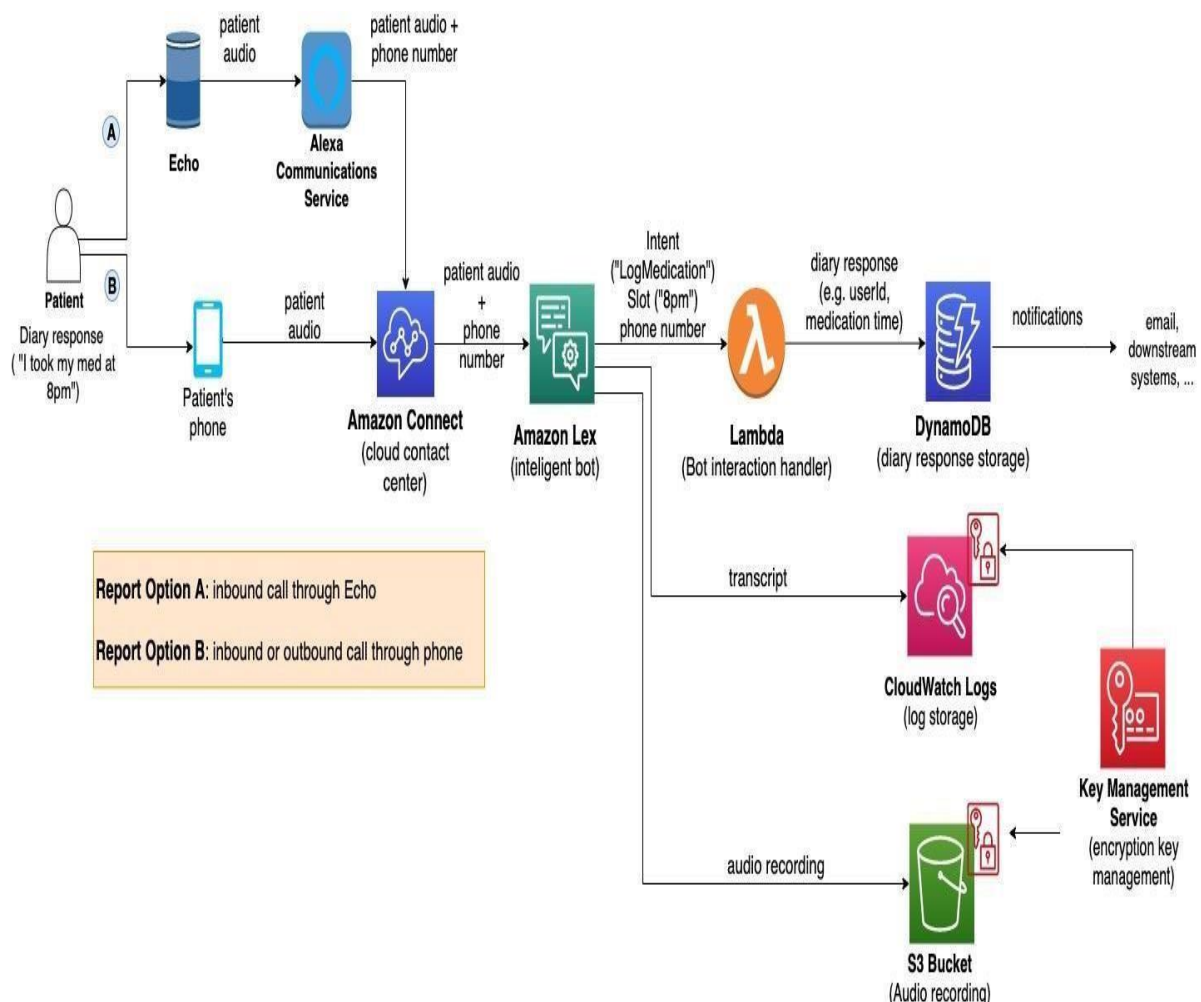
Figure 1: Architecture and data flow of the voice patient diary sample application

5.PROJECT PLANNING & SCHEDULING

5.1 Technical Architecture

The Deliverable shall include the architectural diagram as below and the information as per the table

5.2 Sprint Planning & Estimation

Sprint Schedule, and Estimation

Use the below template to create sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	
Sprint-2		USN-3	As a user, I can register for the application through Facebook	2	Low	
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Medium	
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	
	Dashboard					

6.CODING &SOLUTIONING

6.1 Data Collection & Extraction from Database

Activity 1: Collect the dataset

Please use the link to download the dataset:-

https://www.kaggle.com/datasets/kanishk307/6000-indian-food-recipes-dataset

Activity 1.1: Understand the data

Check the below link out to understand the dataset in detail:-

https://www.kaggle.com/datasets/kanishk307/6000-indian-food-recipes-dataset

Activity 2: Connect IBM DB2 with IBM Cognos Explanation video link:

https://drive.google.com/file/d/1VzWCqOKKUrJrYpPc6TUwYTaF19o6ni5r/view?usp=sharing

Milestone 3: Data Preparation

Activity 1: Prepare the Data for Visualization

Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring the data is accurate and complete. This process helps to make the data easily understandable and ready for creating visualizations to gain insights into the performance and efficiency.

Activity 1.1 : Preparing a Data Module

https://www.kaggle.com/datasets/kanishk307/6000-indian-food-recipesdataset

Milestone 4: Data Visualization

Data visualization is the process of creating graphical representations of data in order to help people understand and explore the information.
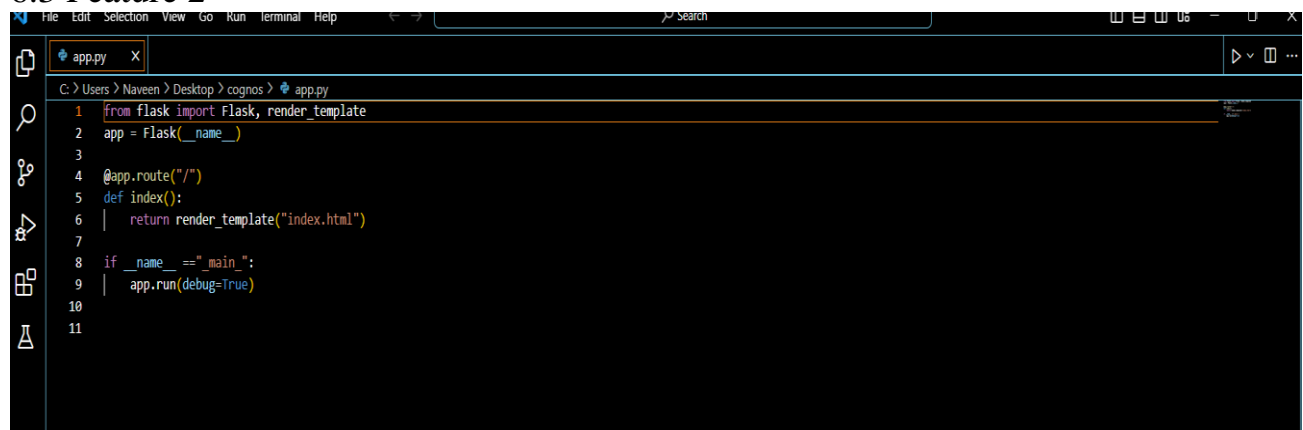
The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

6.2 Feature 1

6.3 Feature 2

7.PERFORMANCE TESTING

The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyze the Kaggle data include bar charts, line charts, heat maps, scatter plots, pie charts, maps etc. These visualizations can be used to compare performance, track changes over time, show distribution, and relationships between variables.

Activity 1.1: Value by Area?
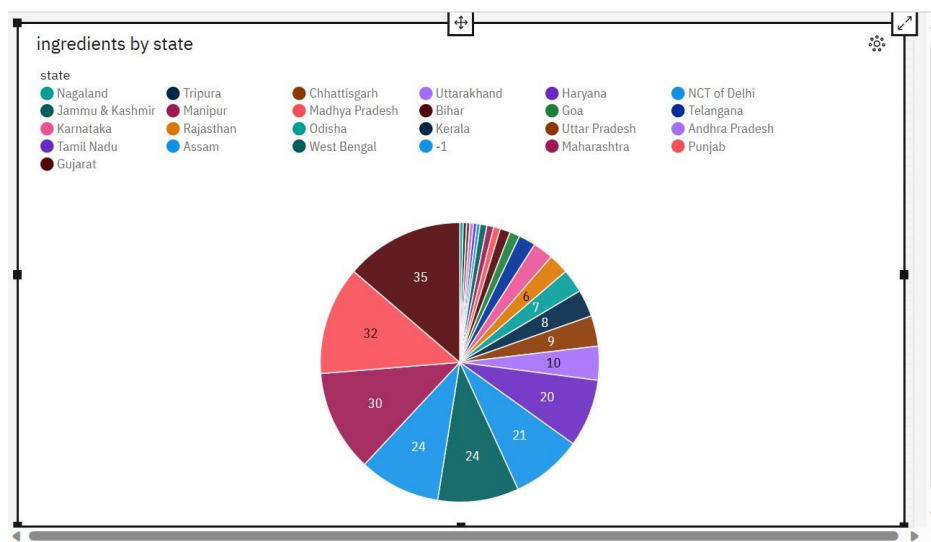
Explanation video link:

Milestone 5: Dashboard

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

Activity 1: Responsive and Design of Dashboard

The Indian Food EDA Dashboard offers a comprehensive and interactive visual representation of the insights derived from exploring a rich dataset on Indian cuisine. This dashboard provides a user-friendly interface that allows users to delve into the diverse flavours, regional variations, popular dishes, and nutritional aspects of Indian food. With a range of interactive visualizations and data exploration tools, the dashboard facilitates a deeper understanding and appreciation of Indian culinary traditions.
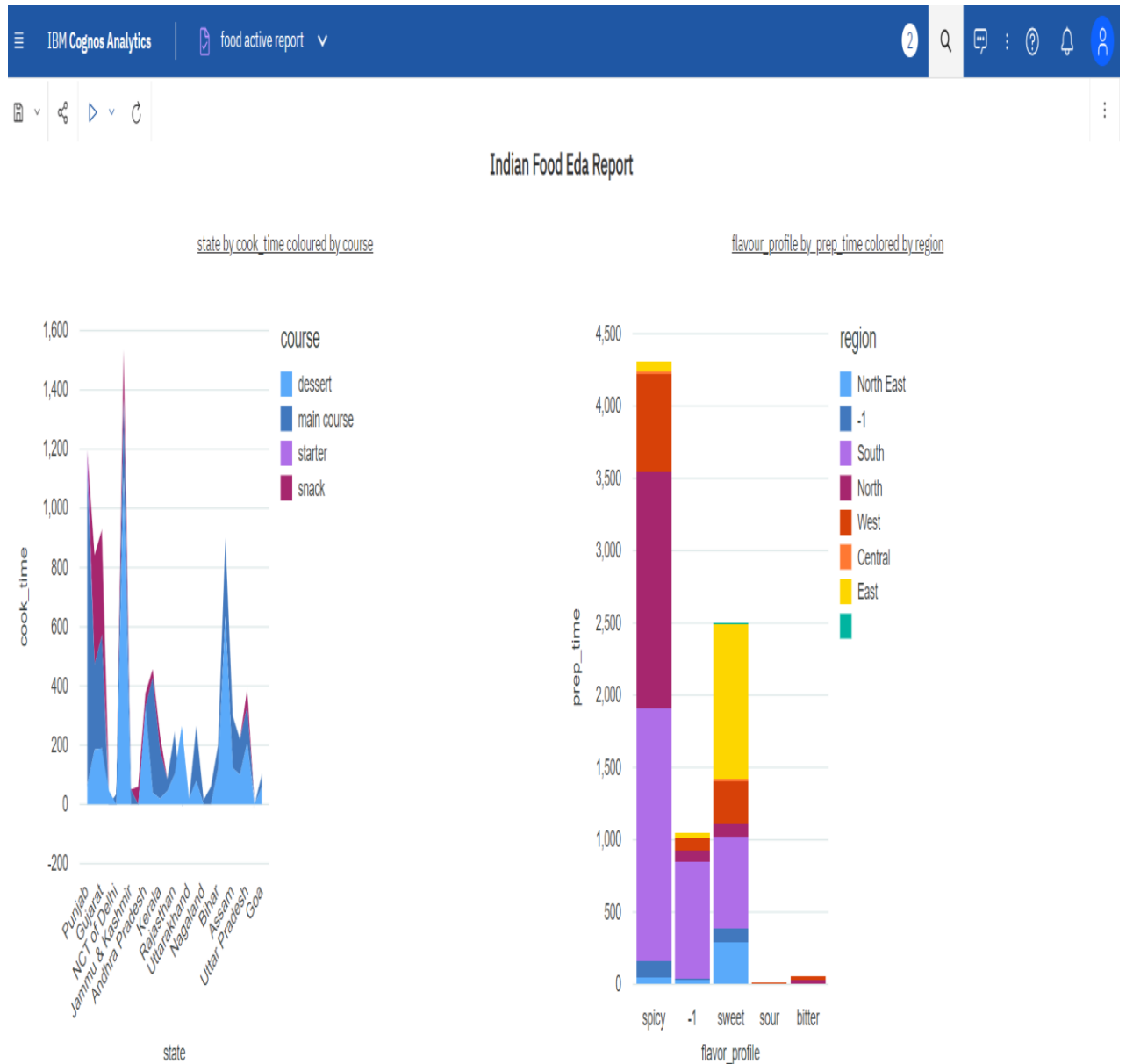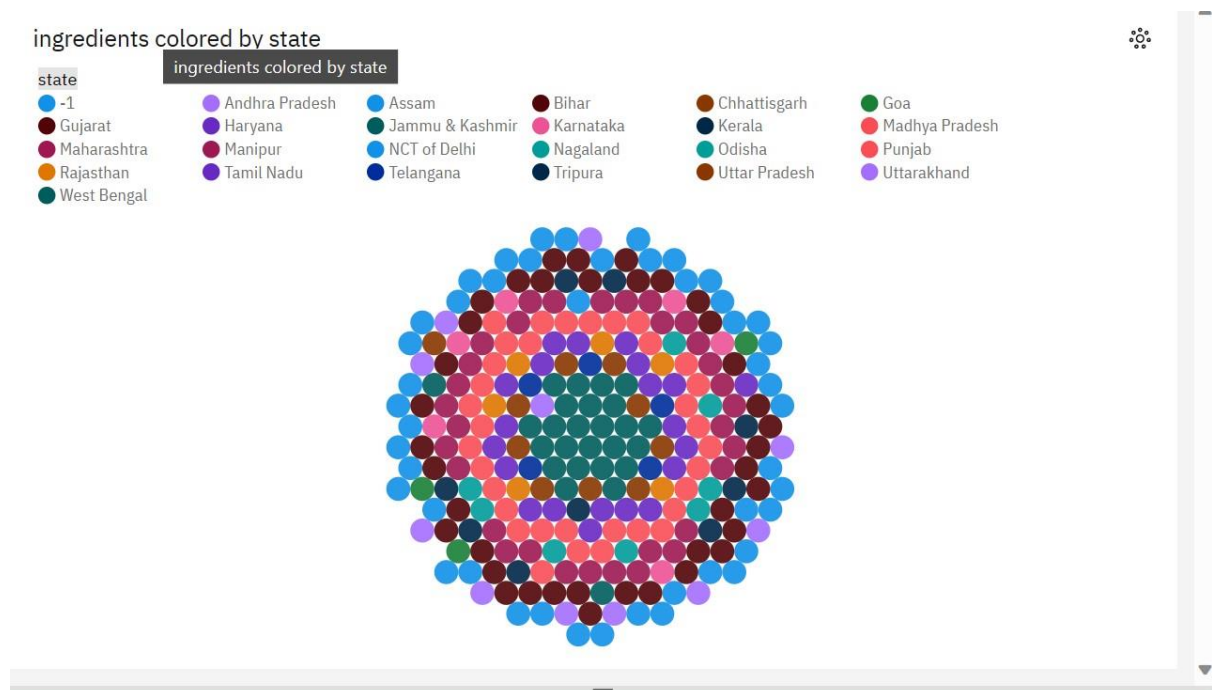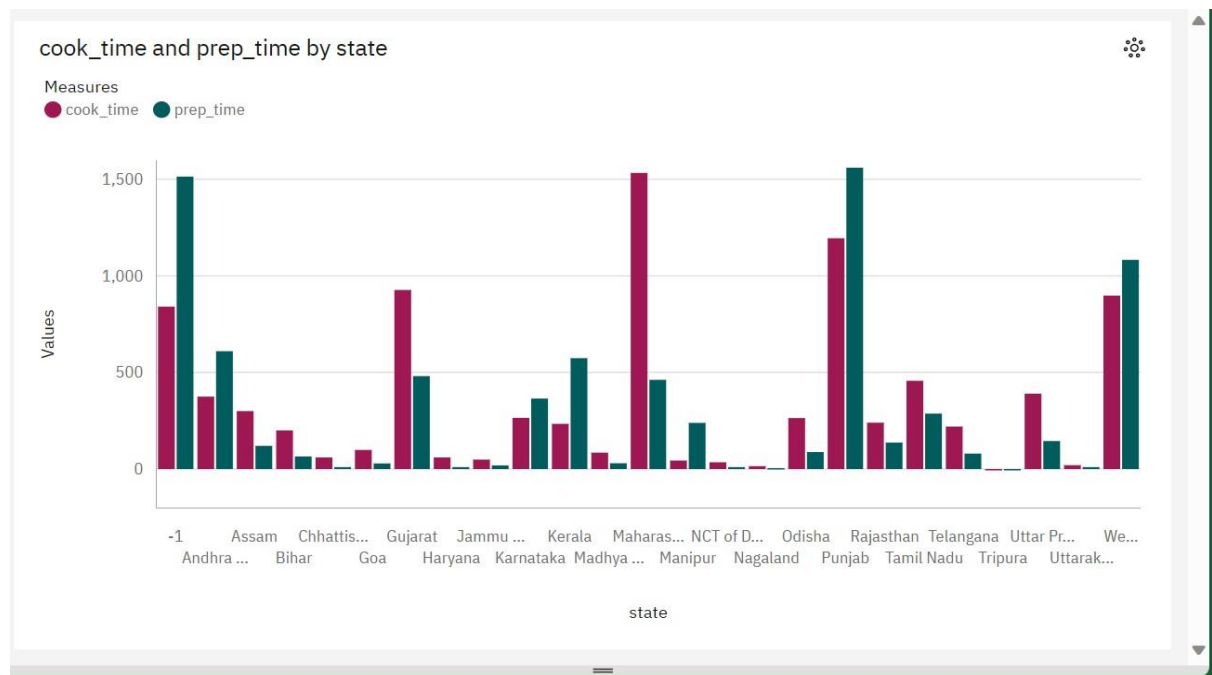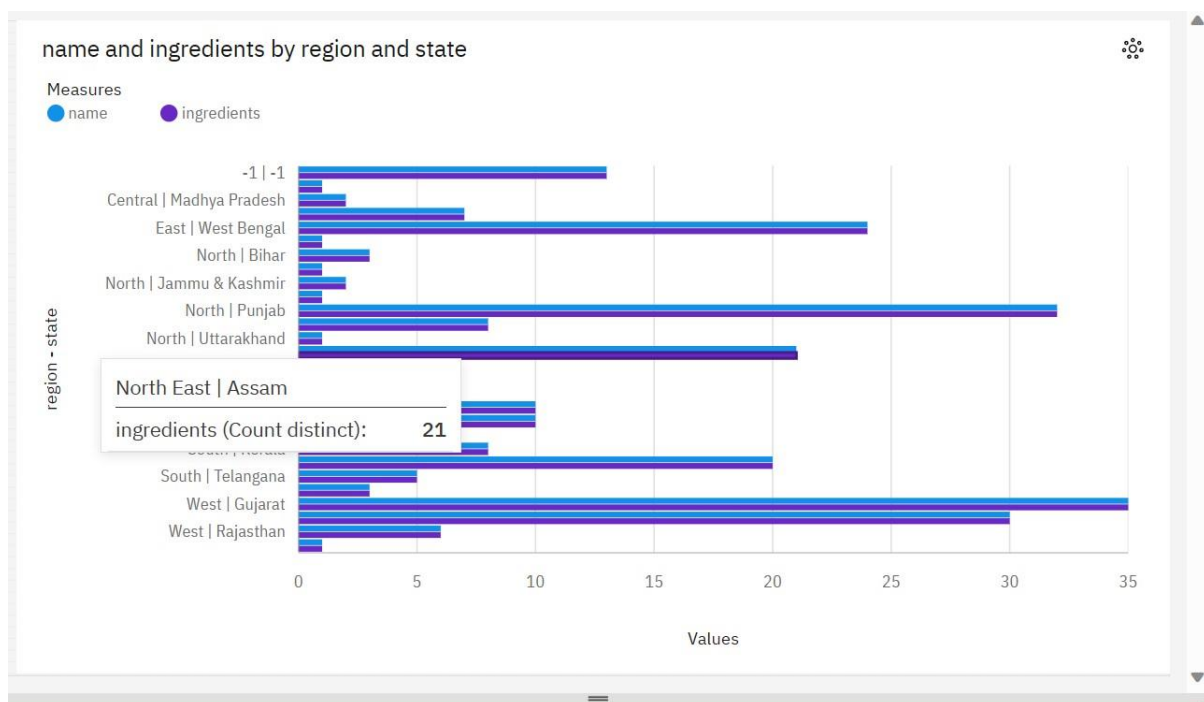
Explanation video link:
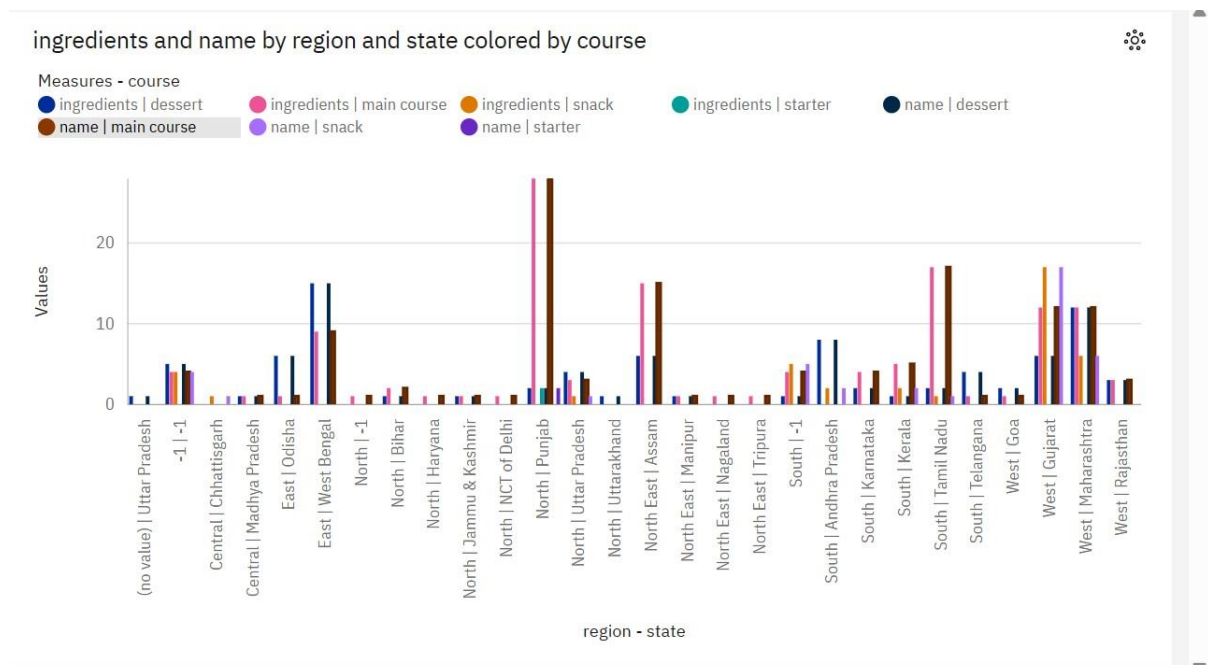
Milestone 6: Report

A report is a document that presents information in a specific format and layout, usually based on data from a database or other data source. A report in IBM Cognos can contain various elements, such as tables, charts, graphs, and images, as well as text and data elements, and it is designed to be used by business users to help them better understand their data and make informed decisions. There are several different types of reports available in IBM Cognos, including list reports, crosstab reports, chart reports, and report studio reports, among others. The type of report that you choose will depend on the specific needs and requirements of your organization, as well as the data that you need to present.

8.RESULTS

cook_time and prep_time by state

Measures
- cook_time
- prep_time

ingredients colored by state

state
- -1
- Andhra Pradesh
- Assam
- Bihar
- Chhattisgarh
- Goa
- Gujarat
- Haryana
- Jammu & Kashmir
- Karnataka
- Kerala
- Madhya Pradesh
- Maharashtra
- Manipur
- NCT of Delhi
- Nagaland
- Odisha
- Punjab
- Rajasthan
- Tamil Nadu
- Telangana
- Tripura
- Uttar Pradesh
- Uttarakhand
- West Bengal

prep_time and cook_time by ingredients

name and ingredients by region and state

Milestone 7: Story

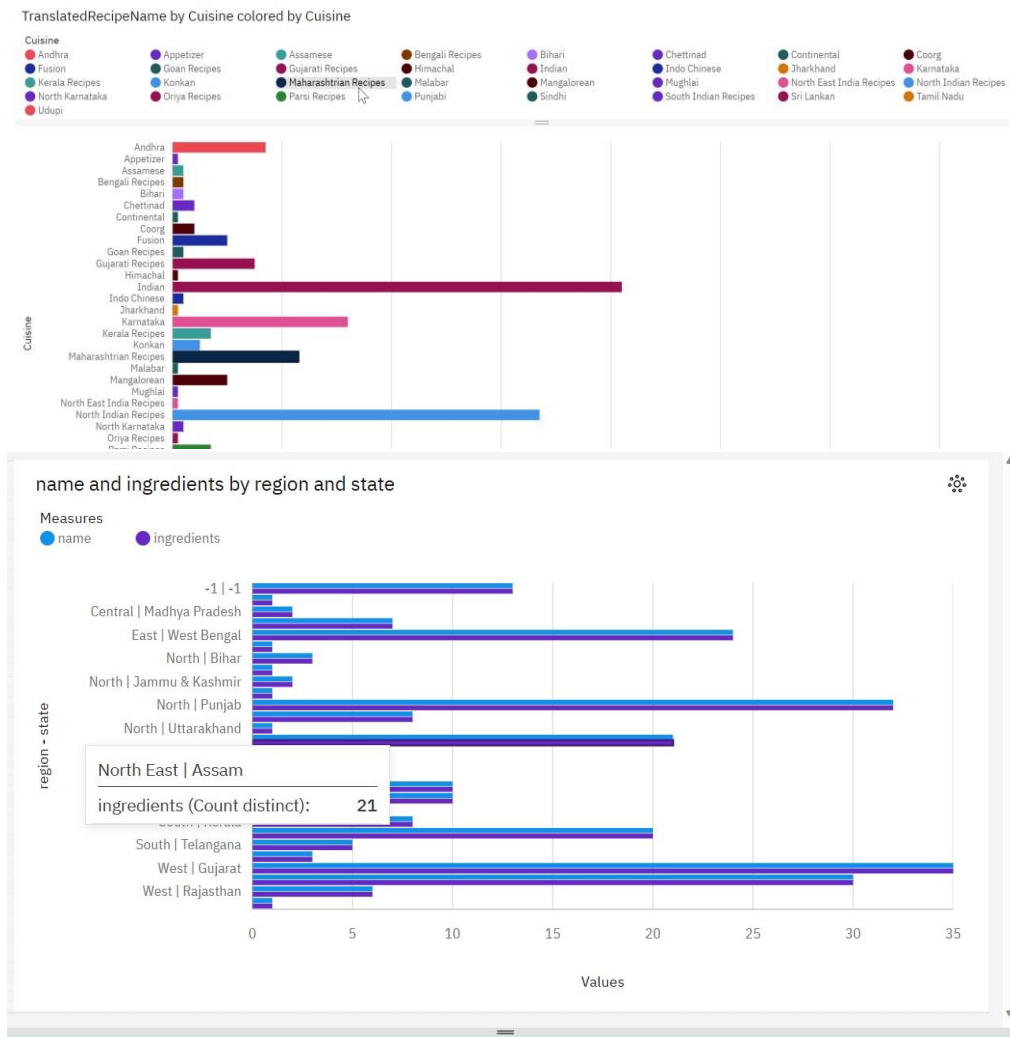A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and

highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos. Activity 1: No of Scenes of Story

Embark on a captivating culinary journey through the rich and diverse flavors of India with our interactive story based on Exploratory Data Analysis (EDA) of Indian food. Immerse yourself in the vibrant world of Indian cuisine as we uncover the secrets hidden within a comprehensive dataset. From savoring regional delights to exploring popular dishes and understanding the nutritional aspects, this story takes you on an unforgettable gastronomic adventure.

Explanation video link: https://drive.google.com/file/d/1l9dlN-BsRgnS3weGFSfknZ5nr2Qbb1E/view?usp=drive_link

TranslatedRecipeName by Cuisine colored by Cuisine

name and ingredients by region and state

Measures
● name   ● ingredients

North East | Assam

ingredients (Count distinct):   21

Milestone 9: Web integration

Publishing helps us to track and monitor key performance metrics, to communicate results and progress. help a publisher stay informed, make better decisions, and communicate their performance to others. Integrating dashboard/reports/stories to web

Step 1: Go to Dashboard/story/report, click on share button on the top ribbon

Note: You can also change the width and height of the dashboard/story/report as you like.

Activity 1: Dashboard and Story embed with UI With Flask

Explanation video link:

Flask Deployment :- * Running on http://127.0.0.1:1200/ (Press CTRL+C to quit)

9.ADVANTAGES

Comprehensive Analytics Platform: IBM Cognos offers a wide range of analytics capabilities, including reporting, dashboarding, data visualization, and advanced analytics, making it a versatile tool for various analytical needs.

Scalability: Cognos can handle large volumes of data and scale to meet the needs of both small businesses and large enterprises.

Integration: It integrates well with various data sources, including databases, data warehouses, and cloud-based data platforms, allowing users to access and analyze data from different sources in one place.

Self-Service Analytics: Cognos provides self-service features, allowing non-technical users to create reports and dashboards without relying on IT support.

Security and Governance: It offers robust security and governance features, helping organizations maintain data integrity and comply with regulations.

AI and Machine Learning: IBM Cognos has AI and machine learning capabilities that can help users discover insights, predict trends, and automate routine task

DISADVANTAGES

Cost: IBM Cognos is known to be relatively expensive, which may be a barrier for small businesses or organizations with budget constraints.

Complexity: Setting up and configuring Cognos can be complex, and it may require significant training for users to take full advantage of its features.

Resource Intensive: Running Cognos efficiently often requires substantial hardware

and IT resources, especially for large deployments.

Learning Curve: Users, especially beginners, may find Cognos challenging to learn

due to its feature-rich nature.

Limited Data Sources: While Cognos supports various data sources, some users may find it less flexible when dealing with unconventional or niche data sources.

Customization: Customizing reports and dashboards to meet specific business needs can sometimes be time-consuming and require technical expertise

10.CONCLUSION

The EDA on Indian food dataset provides valuable insights into the characteristics, variations, and popularity of Indian cuisine. The analysis helps in understanding the diversity of Indian food, identifying key ingredients, regional variations, and popular dishes. These insights can be useful for culinary enthusiasts, food researchers, nutritionists, and even restaurant owners looking to understand Indian cuisine better or develop new recipes.