FlavourFlix

Discover. Taste. Repeat.

Mixing Food with Technology
Proposal for FlavourFlix

Capstone Project – 2023/2024

Group 4

Bruno Moreira – Business Analyst and Developer

Carolina Braziel Shaul – Project Manager and Developer

Guilherme Carriço – Quality Assurance and Developer

Madalena Frango – Marketing Specialist and Developer

Table of Contents

Contextualization Note

The present document describes the definition and respective plan of implementation of the final project selected by Group 4 for the curricular unit Capstone Project.

The principal aim of the further detailed project is to put in action the several technologies and methodologies presented throughout the Data Science bachelor's degree and articulate them in a complementary fashion to develop a user-based and business-oriented platform.

It is expected to incorporate a vast set of techniques and data science frameworks such as data retrieval, preprocessing and analysis, machine learning, text mining, natural language processing and API integration. Additionally, it is desired to articulate tools like Large Language Models, as means of facilitating the design and implementation of the project, whilst offering new perspectives.

Furthermore, this report represents the first deliverable of the afore mentioned project and aims to describe the group's Project Proposal. It is important to highlight that the project is currently in preliminary stages, and thus the ideas further presented may vary throughout its development.

Executive Summary

In order to fulfil the afore mentioned objectives, the group opted to develop a web-based platform and service denoted by "*FlavourFlix"*, a business within the food and restauration technology industry catered to food enthusiasts.

The rationale is to integrate the social network elements present in food and restauration-related platforms such as *TheFork* and *Zomato* combined with the recommendation capabilities that distinguish *Netflix* and *YouTube*, by enabling an end-user to not only discover what restaurants are closest to them and their respective characteristics, but also to receive restaurant suggestions based on a series of prompts provided to a chatbot Filomena by the user. To complement, the latter will be warranted the possibility of simulating reservations grounded on logistic related factors such as current table availability. Moreover, Filomena will be able to, while discussing recommendations with the user, decipher which kind of food-personality category they best fit in.

1. Context

For many, dining-out represents a privilege. Furthermore, Portugal is a country notably internationally acknowledged for its traditional cuisine value and hospitality-oriented service. When you pair these factors togethers, it becomes only natural for someone to desire to experience the best of the best.

In addition, it is currently estimated that food-technology and travel assistance services such as *TheFork* and *TripAdvisor* have been contributing in a significantly positive manner to the global growth of the restaurant and hospitality industry.

2. Problem Definition

However, the amount of restaurants and their variety, allied to the personal preferences and expectations of each person, typically lead to a sense of indecisiveness. Nonetheless, as tourism expands, and the restauration industry evolves jointly, certain typical meals become harder to find without a thorough search between friends or online, which can become an effortful task. The ever-increasing range and availability of information surrounding us proves the need of finding adequate information retrieval systems, which aim to narrow down searches while maintaining their substance.

Moreover, as the industry of personalized products and services continue to trend, users become more demanding and interested in customized experiences that highlight their specific needs and desires.

3. FlavourFlix's Mission

This is where FlavourFlix comes to play. The situations described prior are some of the practical illustrations of possible problems the business aims to tackle, by providing with ways of ensuring customers are able to explore the cuisine in Portugal in a straightforward and trouble-free fashion, while extracting the most out of their discoveries. Thus, the purpose with FlavourFlix is to facilitate those who are passionate about cuisine experiences in finding curated sets of restaurants in Portugal given their explicit requests.

4. Meet FlavourFlix

"Introducing FlavourFlix, your passport to culinary adventure! Are you tired of the same dining routine, or feel like ditching the kitchen and indulge your taste buds? Whether you're an adventurous eater or just looking for a fresh experience, Flavour Flix is your new trusted pattern to discover hidden cuisine gems around you."

4.1. The Company

4.1.1. Vision and Values

FlavourFlix is a business which aims to provide quality service to its customers and is mainly characterized by a web-platform designed to facilitate consumers' decision-making and expand their dining experience. It is mainly based on data regarding menus, ratings, photos, cuisine, style, pricing, and other characterizing factors of a wide set of restaurants located in Portugal, ranging from typical traditional food to exquisite cuisine.

4.1.2. Who is FlavourFlix for?

FlavourFlix is for anyone who appreciates food and dining-out, which – according to our humble estimates – translates into practically anyone. This being said, FlavourFlix caters to an extensive spectrum of individuals who share an appreciation for culinary and dining experiences. The app's target audience is multifaceted, spanning through the younger generations to the most experienced customers. Essentially, FlavourFlix has per objective to always provide a solution - whether you are someone seeking to diversify your palette, a family in pursuit of family-friendly establishments, a traveller looking for authentic Portuguese traditional cuisine or just a picky eater.

4.2. The Product

4.2.1. Analytics Dashboard: Exploring restaurants has never been easier.

As the end-user is first introduced to the FlavourFlix platform, they are redirected to a landing page composed of an **Analytics dashboard** in such a manner that they are able to grasp which nearby restaurants are currently available and serving.

This section of the platform also provides with the possibility of obtaining relevant statistical measures such as the top-rated nearby restaurants, average price distribution, most common types of cuisines or even travelling distance estimations. Essentially, this component intends

for the dynamic exploration of the available data in an intuitive and enticing manner, enabling the end-user to either access information without providing any input, or even narrow down the search by selecting dates and hours, define locations and share preferences in terms of cuisine, menu, chef name, and others.

4.2.2. Meet Filomena, your AI Assistant.

If this proves insufficient to satiate the user's curiosity, or the latter is looking for a different degree of spontaneity, they are able to navigate to another section of the page, where they will meet Filomena. **Filomena** is **FlavourFlix's assistant**, who will, based on some interaction with the user, curate the most appropriate **restaurant suggestions** and ensure they extract the maximum potential from FlavourFlix, consequently ending the day with pleased taste buds and a new experience to share with others.

To finetune Filomena's recommendations, the user may also access the **Profile** section, where they are free to grant relevant information about them such as given allergies, preferred cuisine types, locations, and other factors. By introducing data into this section, Filomena will also be able to **decipher** which "**flavour-personality**" the user possesses.

4.2.3. Can't wait to eat? Reserve now!

As the user explores the different available restaurant options, they are prompted with the possibility of **simulating a reservation** to dine in the space they are currently looking into with Filomena. Additionally, the user will also receive access to the respective restaurant *TheFork* page link or phone number, to perform an actual reservation.

5. Technical Project Description

The rationale and organization of the several technical components of FlavourFlix which have been physically elaborated or at least outlined as to the current date are described further. Note that, as this is a project in progress, some features beyond the further mentioned may constitute the final version of FlavourFlix, as well as some features may be implemented in slightly distinct manners, according to how the project evolves.

5.1. Data Retrieval and Creation

For the implementation of the reported features, it is required to access data regarding several dimensions of each restaurant. Nonetheless, it is also required to possess a client database. To ensure this task, two distinct major types of data were extracted:

- **External data**: Refers to data on real information from restaurants, extracted using APIs designed for web scraping. Specifically, *TheFork* Scraper by *Apify* was utilized to extract from *TheFork* basic information regarding restaurants based on a pre-defined location, being ran for several municipalities of Portugal.

  Within this framework, it was possible to access multiple restaurants scattered throughout the country and extended information on their addresses, phone numbers, *TheFork* user's reviews and ratings, photos, cuisine, style, tags, chefs, offers and restaurant party size.

  This approach revealed unable to gather information on menu data, which consists of crucial information for this task, so it was complemented by the usage of *scraperAPI*, which, based on the restaurant's URLs outputted by *Apify*, was able to retrieve the HTML code associated to each restaurant menu page in order to later extract the menus through text mining techniques.

- **Artificially generated data**: As certain pieces of data were centric to *TheFork's* service, namely the features regarding offers, these were artificially generated based on the remaining restaurant variables such as the restaurant's schedule. Additionally, to extend the capabilities of the project, data regarding the number of currently available tables within each restaurant are simulated each time the user accesses the platform. To complement, data regarding whether the restaurant possesses an outside terrace was also simulated in a random manner.

  The client-centric database was also artificially generated utilizing the AI-based *Mockaroo* platform and further treated for inconsistencies and biases. To define the features which should constitute this database and further prompt *Mockaroo* to populate them, the team decided to only gather information on data that we believe an end-user would be willing to disclose given the context of FlavourFlix, such as age and food preferences in detriment of income and professional activity, for example.

The described data is all stored in *.csv* files within the repository. As this consists of a relatively small project for now, the size of the data is manageable and as a consequence we assume as hypothesis that the information regarding restaurants and clients will be easily accessible from the remaining app features without requiring database systems external to our project repository. The summary of the dataset features, and respective metadata is present in Annex I.

5.2. Web Presentation

The visual implementation of FlavourFlix will be based in Python through the package **Streamlit** and its extensions. For inspiration and guidance, the design of the platform was outlined resulting in the following user journey map:

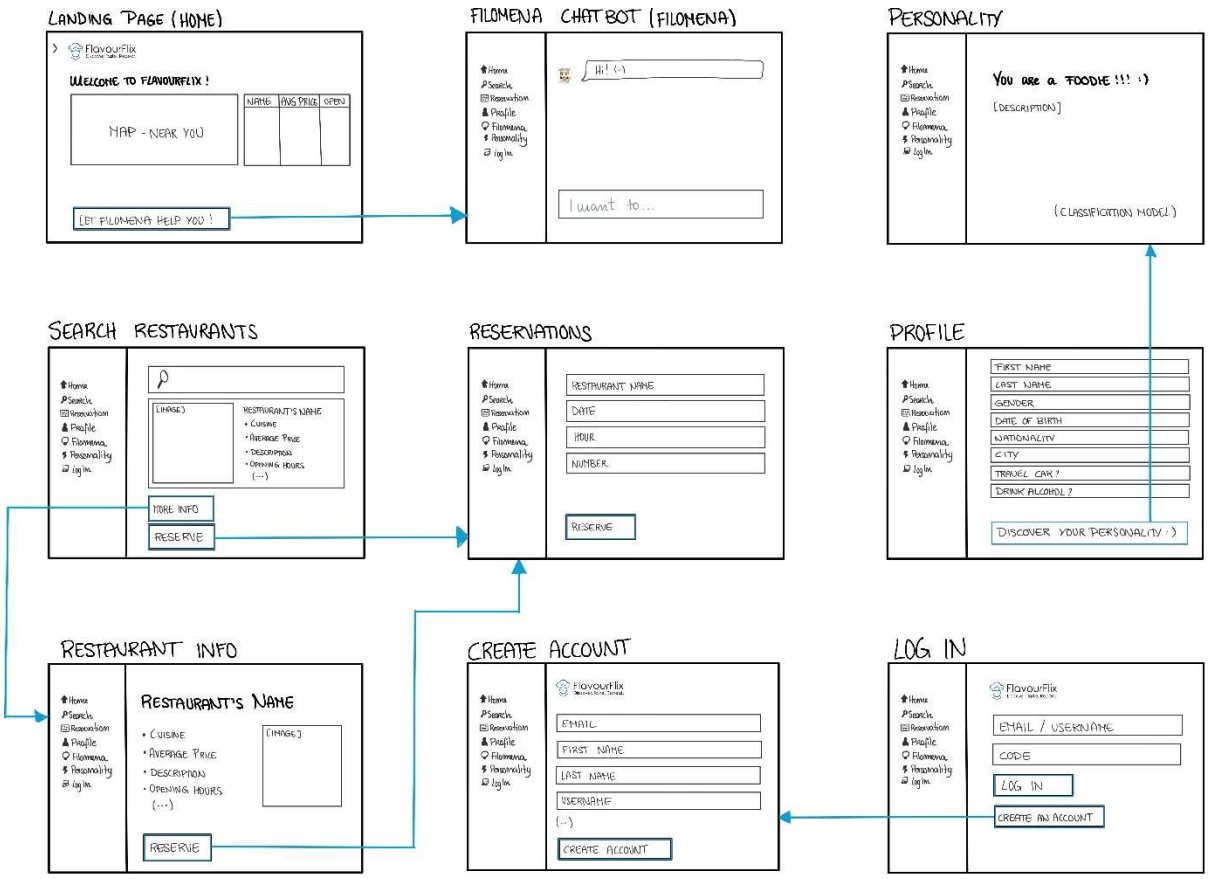

Figure 1: *Outline of FlavourFlix's web platform design.*

5.3. App Features Description and Implementation Proposal

5.3.1. Analytics Dashboard

The analytics dashboard consists of the landing page of the FlavourFlix platform. When accessing the platform, the user will be redirected to this section, where they will be able to gather information about the surrounding available restaurants and obtain some of their general metrics such as average price, most popular cuisine, and others, based on filters provided by the user. We propose to implement this section using simple data exploration techniques through the Python *pandas* package articulated with *streamlit's* capabilities and visualization tools such as *folium*, *seaborn* and *matplotlib*. Additionally, the user will be able to provide their exact location which will narrow down the statistics presented within this section.

5.3.2. Search System

To ensure FlavourFlix's mission is completed, the user will have to be able to easily discover new restaurants based on simple searches. This may be achieved through the **Search Page** of the application. Within this search, the user will either state keywords such as the name of a chef they aim to get to know, a location or even a food they have been craving.

As a consequence, this text input will be fed as a vector into our system in a manner such that the latter will be able to look for close matches within the standardized restaurant information, returning the results that appear most promising based on the user's prompt. The user will then be able to access the pages of the returned observations, and see more information regarding the restaurant and its menu.

5.3.3. Filomena's Restaurant Recommendations

The user will also have, as mentioned before, the opportunity to talk to FlavourFlix's assistant Filomena. Filomena is an AI chatbot capable of finding out which is the most appropriate dining experience for the user, considering a set of messages sent by the latter. As the user introduces a series of prompts, mentioning what they desire to eat, whether they are feeling adventurous or indulgent or what they are allergic to, Filomena is able determine which are the best cuisine suggestions for the user in question. The inputted text is stored as a vector which will be fed into a database used to generate recommendations.

5.3.4. Food-Personality Classification

FlavourFlix's assistant Filomena will be able to, not only provide with restaurant suggestions to the end user as described afore, but also assign each customer to a "Food Personality" based on the information they provide to the platform. This feature aims to showcase our qualities in terms of the two main "sections" of machine learning: unsupervised learning and supervised learning.

Essentially, the task of customer segmentation (unsupervised learning) will be implemented on the simulated client database, in order to determine which features most characterize the customers. By understanding how the customers are alike from each other and through what perspectives they most differ, it is possible to label them with a descriptive "personality trait".

After labelling the customers food tendencies, it is possible to train a supervised learning multi-classification model using their "food-personalities" as the target. Furthermore, once the trained model achieves a certain degree of accuracy, it is intended to use it to classify new

users. The latter will be another of Filomena's tasks, who will be also responsible for enticing the user in discovering this dimension of their personality. For this to be possible, the user must insert their information within the **Profile** section or be willing to disclose it directly to Filomena.

It is important to highlight that the team recognizes that these models may not be entirely unflawed, due to the source of the client data. Since the data is simulated, it is a possibility that no extremely relevant patterns may be found, capable of producing highly differencing labels.

5.3.5. Recognizing the Customer

The user will be able to access the aforementioned services by creating an account with the FlavourFlix platform. To do so, it is only required a valid e-mail and username, which will be stored in the respective database.

After the account is successfully created, the user will be able to log-in to the platform with either any of the defined credentials, and will be prompted to enter an entry code which will be sent by FlavourFlix. This authentication mechanism was introduced due to privacy safeguards. As to the date the team has not yet guaranteed a completely private and secure procedure in the way the data is stored. Thus, the team opted to develop a password-free mechanism to avoid any potential hassles of sensitive information.

Through these authentication steps our system is able to recognize the user and finetune its functionalities according to their preferences.

5.4. Privacy Concerns

Within the context of this project, data must be extracted and utilized from multiple sources, making it essential to address the topic of privacy, in order to ensure responsible handling.

Throughout the implementation of this task, it will be utilized publicly available data (e.g. restaurants data, primarily retrieved from TheFork website), with all associated rights recognized. Additionally, it is going to be utilized private data (e.g. users' data, whether fictitious or non-fictitious, with explicit user consent in the last case).

This data will be used and stored only until the project's end date, and it is exclusively intended for academic purposes, forming the basis for developing the proposed tools within the context of the curricular unit Capstone Project of the 3rd year of the bachelor's degree in Data Science at NOVA Information Management School (NOVA IMS).

Given the academic nature of the project, all individuals with access to the project's resources will have access to the mentioned data.

5.5. Roadmap: Project Tasks Timeline and Milestones

In order to organize and further achieve the proposed components of FlavourFlix, the following roadmap was defined, based on the widely used methodology CRISP-DM (Cross Industry Standard Process for Data Mining), thus ensuring a structured and systematic approach to the problem at hands.

Milestone 1: Project Inception and Team Formation (Week 1)

- Form the project team, assign roles and responsibilities.
- Define the project scope, objectives, and key deliverables.
- Create a detailed project plan, including timelines and resources.
- Conduct initial research on data sources and data retrieval techniques.

Milestone 2: Data Gathering and Preprocessing (Weeks 2 - 8)

- Collect restaurant data from various sources (e.g., *TheFork*).
- Simulate client data through sources such as *Mockaroo* and further normalization to prevent biases.
- Develop data preprocessing pipelines to clean and structure the data.
- Ensure data quality, consistency and accuracy through data validation and cleaning procedures.
- Feature engineering.

Milestone 3: User Interface Development (Weeks 8 - 15)

- Design the user interface using the Streamlit interface.
- Implement features for displaying nearby restaurants, statistics, and filtering options.
- Test and validate the functionality of the user interface to guarantee quality aesthetics, functionality, and accessibility for all.

Milestone 4: Filomena Chatbot Development (Weeks 8 - (undefined))

- Design and develop the Recommendation system.
- Implement the Recommendation System in an articulated fashion with a Chatbot.

Milestone 5: Customer Personality Classification (Weeks 8 – 13)

- Perform unsupervised learning techniques on the current customers data to label their food-personalities.
- Develop a supervised learning classification algorithm to predict which food-personality the customer belongs to, given the information they provided.

5.6. Evaluating Success

With the objectives and key results clearly identified, it is important to define a way to measure the success of this project. Based on some frameworks widely utilized and the academic context in which this project emerges, the following is proposed:

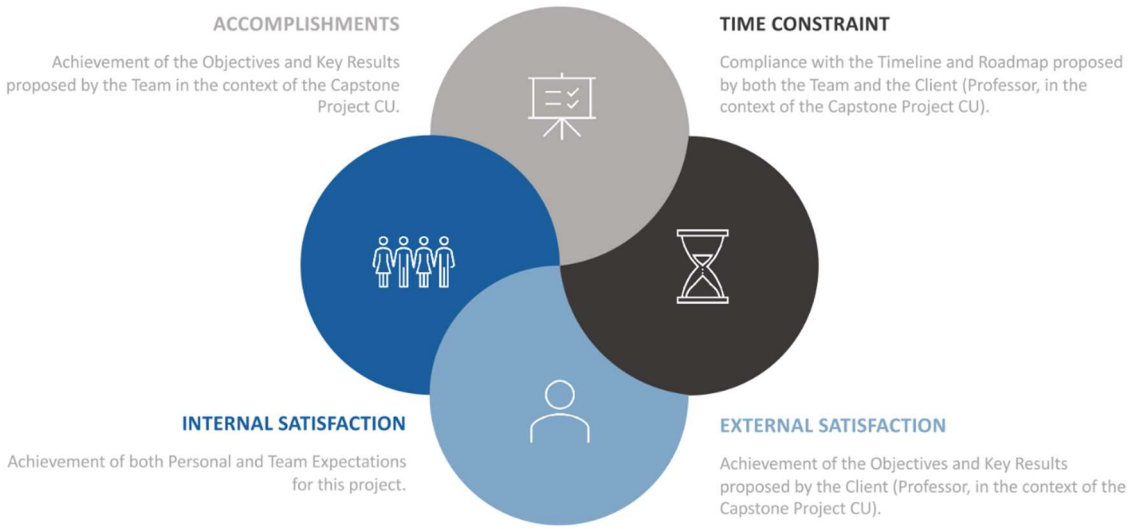

Figure 2: *Key-factors to measure the success of FlavourFlix.*

As already mentioned, the present document is only an initial proposal and, for this reason, the project is subject to any modifications in any of its components.

Annexes

Annex I – Project Features and Description.

***Table 1**: Summary of restaurant data and metadata.*

Feature Name	Feature Description	Feature Name	Feature Description
url	Restaurants *TheFork's* url.	*phone*	Restaurant's phone number.
name	Name of the restaurant.	*ratingValue*	Restaurant's average rating value in *TheFork*.
address	Address of the restaurant.	*reviewCount*	Number of reviews within *TheFork*.
photo	First picture in *TheFork's* restaurant page.	*style*	Restaurant's style category.
averagePrice	Average price per person in euros of the restaurant.	*latitude*	Latitude coordinate of the restaurant's address.
chefName1	Restaurant's chef name (1).	*longitude*	Longitude coordinate of the restaurant's address.
chefName2	Restaurant's chef name (2).	*location*	Restaurant's location value.
chefName3	Restaurant's chef name (3).	*city*	City the restaurant is located in.
cuisine	Type of cuisine of the restaurant.	*ambienceRatingsummary*	Average rating value in terms of ambience.
michelin	Whether the restaurant has a Michelin star.	*foodRatingSummary*	Average rating value in terms of food quality.
description	*TheFork's* description of the restaurant.	*serviceRatingSummary*	Average rating value in terms of service.
isBookable	Whether the restaurant has the ability to book a table.	*paymentAcceptedSummary*	Possible payment mechanisms accepted by the restaurant.
maxPartySize	Maximum number of people allowed per table.	*menu_pre_proc*	Restaurant's menu.
schedule	Restaurant's schedule.	*currentCapacity*	How full is the restaurant currently.
promotions	Restaurant's possible promotions.	*outdoor*	Whether the restaurant has outdoor eating section or not.

***Table 2**: Summary of client's data and metadata.*

Feature Name	Feature Description	Feature Name	Feature Description
first_name	Customer's first name.	*smoker*	Whether the customer smokes.
last_name	Customer's last name,	*allergies*	Which food allergies the customer has.
email	Customer's platform email.	*favourite_food*	Customer's preferred food or meal.
gender	Customer's gender.	*dislike_food*	Customer's most disliked food or meal.
username	Username associated to the platform.	*preferred_payment*	Customer's preferred payment method.
date_of_birth	Customer's date of birth.	*restaurant_style*	Customer's preferred restaurant style.
nationality	Customer's nationality.	*cuisine_type*	Customer's preferred type of cuisine.
city	District the customer is currently based on.	*lunch_hour*	Customer's typical lunch hour range.
travel_car	Whether the customer travels by car.	*dinner_hour*	Customer's typical dinner hour range.
drinks_alcohol	Whether the customer drinks alcohol.	*price_range*	Average price the customer is willing to pay per meal per person in euros.
dietary_restrictions	Which dietary restrictions the customer has.		