**Zomato Restaurant Clustering and Sentiment Analysis**

**Introduction**

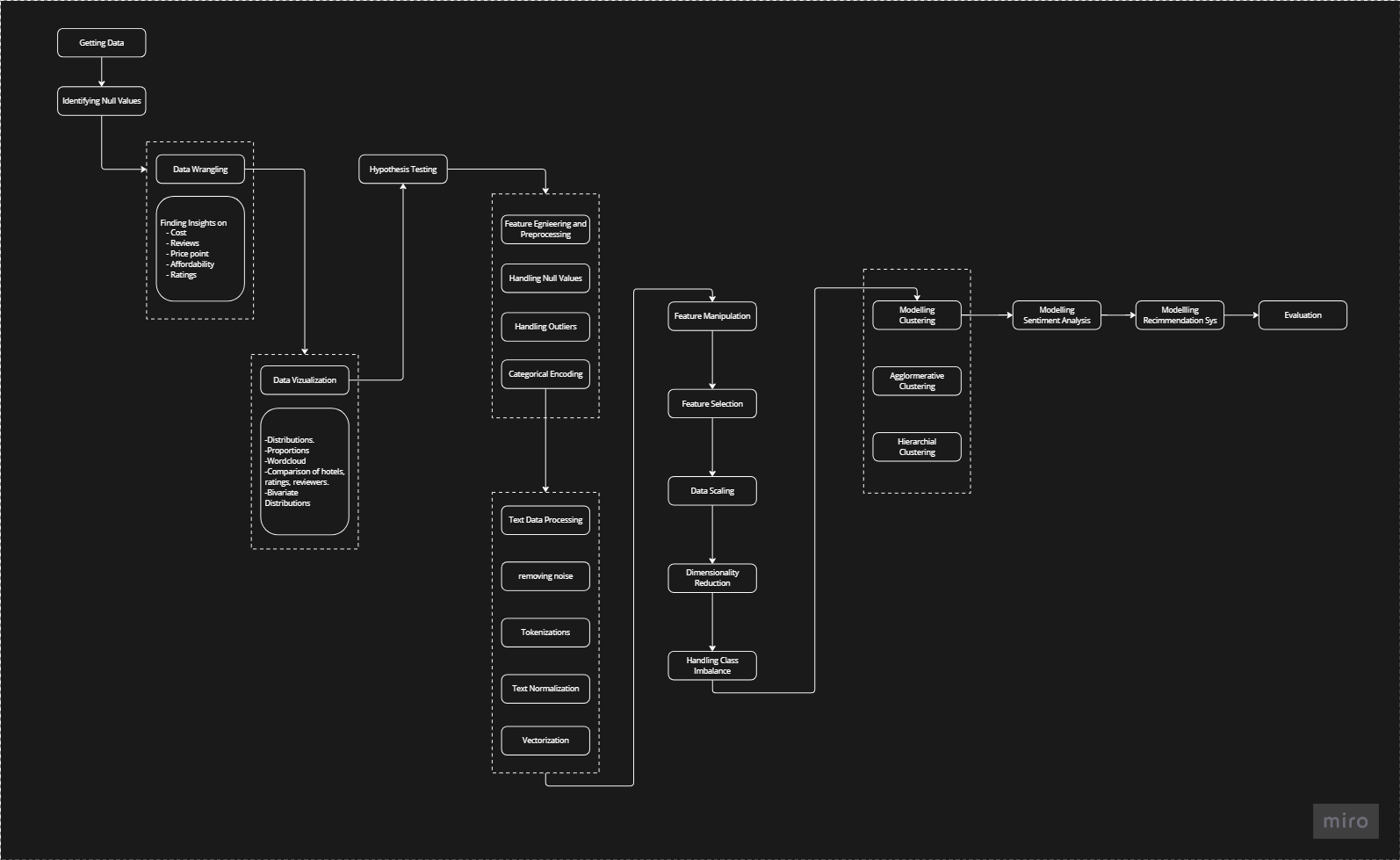
Zomato, founded by Deepinder Goyal and Pankaj Chaddah in 2008, is an Indian restaurant aggregator and food delivery startup. It provides information, menus, and user reviews of restaurants, and offers food delivery options from partner restaurants in select cities. This project focuses on analyzing Zomato restaurant data for various cities in India to gain insights, identify interesting facts and figures about the Indian food industry, and provide useful conclusions in the form of visualizations.

**Problem Statement**

The project aims to analyze customer sentiments from reviews and cluster Zomato restaurants into different segments. This analysis will help customers find the best restaurants in their locality and assist the company in identifying areas for improvement. Additionally, the data provides valuable information on cuisine and costs, which can be used for cost vs. benefit analysis.

**Data Description**

The project uses two main datasets:

1. **Zomato Restaurant Metadata:**
   * **Name**: Name of the restaurant.
   * **Links**: URL links of the restaurants.
   * **Cost**: Estimated cost per person for dining.
   * **Collection**: Tagging of restaurants with respect to Zomato categories.
   * **Cuisines**: Cuisines served by the restaurants.
   * **Timings**: Restaurant timings.
2. **Zomato Restaurant Reviews:**
   * **Restaurant**: Name of the restaurant.
   * **Reviewer**: Name of the reviewer.
   * **Review**: Review text.
   * **Rating**: Rating provided by the reviewer.
   * **MetaData**: Reviewer metadata (number of reviews and followers).
   * **Time**: Date and time of the review.
   * **Pictures**: Number of pictures posted with the review.  
       
     

**Notebook Breakdown**

**1. Business Problem Analysis**

* Understanding the problem and defining objectives.

**2. Data Collection**

* Gathering data from Zomato restaurant metadata and reviews.

**3. Data Cleaning and Preprocessing**

* Handling missing values, data type conversions, and merging datasets.

**4. Feature Engineering**

* Creating new features from existing data to enhance model performance.

**5. Exploratory Data Analysis (EDA)**

* Visualizing data to uncover patterns and insights.

**6. Best Restaurants in the City**

* Identifying top-rated restaurants in various cities based on reviews and ratings.

**7. The Most Popular Cuisines in Hyderabad**

* Analyzing cuisine preferences and trends in Hyderabad.

**8. Restaurants and their Costs**

* Examining the cost distribution of restaurants.

**9. Cost-Benefit Analysis**

* Evaluating the cost vs. benefits of dining at different restaurants.

**10. Hypotheses Generation on Visualized Data for Clustering**

* Formulating hypotheses based on visualized data for clustering analysis.

**Restaurant Clustering**

**11. K-Means Clustering on Cost and Ratings**

* Clustering restaurants based on cost and ratings using the K-Means algorithm.

**12. Multi-Dimensional K-Means Restaurant Clustering**

* Performing clustering on multiple features to identify distinct restaurant segments.

**13. Principal Component Analysis**

* Reducing dimensionality for clustering analysis.

**14. Silhouette Score**

* Evaluating the quality of clustering using the silhouette score.

**15. Cluster Exploration**

* Exploring the characteristics of each cluster to derive meaningful insights.

**Sentiment Analysis**

**16. Exploratory Data Analysis**

* Analyzing the distribution of sentiments in the review texts.

**17. Critics in the Industry**

* Identifying influential reviewers and critics based on metadata.

**18. Text Pre-Processing and Text Visualization**

* Preprocessing review texts and visualizing sentiment distribution.

**19. Modeling**

* Building and evaluating sentiment analysis models.

**Conclusion**

The project successfully analyzed Zomato restaurant data to gain insights into customer sentiments and restaurant segmentation. Key findings include:

* Identification of top-rated restaurants and popular cuisines in various cities.
* Cost-benefit analysis of dining options.
* Clustering of restaurants into distinct segments based on cost, ratings, and other features.
* Sentiment analysis of customer reviews to understand customer satisfaction and areas for improvement.

**Future Work**

* Further refinement of clustering models using advanced techniques.
* Incorporation of additional features such as location and demographic data for more precise analysis.
* Continuous model evaluation and updating with new data to maintain accuracy and relevance.