**Data Collection Plan & Raw Data Sources Identification Report**

1.1 Project Objective:

To collect high-quality and relevant data for the development of a robust restaurant recommendation system. This data will be crucial for training and evaluating machine learning models that accurately predict user preferences and provide personalized recommendations.

1.2 Data Requirements:

User Data:

Demographics: Age, gender, location, occupation, income level

Preferences: Cuisine preferences (e.g., Italian, Mexican, Indian), dietary restrictions (e.g., vegetarian, vegan, gluten-free), price range, dining occasion (e.g., casual, fine dining, romantic), preferred ambiance (e.g., quiet, lively, family-friendly)

Past Behavior: Historical dining data (restaurants visited, ratings, reviews, check-ins), browsing history, search queries

Social Media Activity: Food-related posts, interactions (likes, comments, shares) on social media platforms

Restaurant Data:

Location: Address, coordinates (latitude, longitude)

Cuisine: Type of cuisine (e.g., Italian, Mexican, Indian), sub-cuisine (e.g., Neapolitan, Tex-Mex, South Indian)

Attributes: Ambiance (e.g., casual, fine dining, romantic), price range, service type (e.g., dine-in, takeout, delivery), special features (e.g., outdoor seating, live music, kids' menu)

Reviews and Ratings: User reviews, ratings from various platforms (e.g., Google Maps, Yelp, TripAdvisor)

Social Media Presence: Social media accounts, engagement metrics (e.g., number of followers, likes, shares)

Contextual Data:

Location Data: Current user location, nearby points of interest (e.g., landmarks, attractions)

Time Data: Time of day, day of the week, special occasions (e.g., holidays, weekends)

Weather Data: Current weather conditions (e.g., temperature, precipitation)

1.3 Data Sources:

Primary Data:

User Surveys/Questionnaires: Collect user demographics, preferences, and feedback through online surveys or in-app questionnaires.

User Interactions: Track user interactions with the recommendation system (e.g., clicks, views, selections)

Secondary Data:

Public Datasets: Utilize publicly available datasets (e.g., OpenStreetMap, Foursquare) for restaurant locations, attributes, and user reviews.

Social Media APIs: Access user data and restaurant information from social media platforms (e.g., Facebook, Instagram, Twitter).

Restaurant Review Platforms: Collect data from online review platforms (e.g., Yelp, TripAdvisor, Google Maps).

Third-party Data Providers: Utilize data providers specializing in location-based services, consumer behavior, and demographic information.

1.4 Data Collection Methods:

APIs: Utilize APIs to programmatically access data from various sources (e.g., social media platforms, map services).

Web Scraping: Extract data from websites and online platforms using web scraping techniques (with ethical considerations and respecting website terms of service).

Data Aggregation: Combine data from multiple sources to create a comprehensive dataset.

1.5 Data Storage and Management:

Data Warehouse: Establish a secure and scalable data warehouse to store and manage the collected data.

Data Security: Implement appropriate security measures to protect user data privacy and comply with relevant regulations (e.g., GDPR, CCPA).

1.6 Data Quality Considerations:

Data Accuracy: Ensure data accuracy and completeness through data validation, cleaning, and deduplication.

Data Consistency: Maintain consistency across different data sources and formats.

Data Bias: Identify and mitigate potential biases in the data (e.g., selection bias, representation bias).