**Sampling Techniques Project**

**Problem Statement**

What are the meal preferences for the Innscor food chain across the different regions in Kenya?

1. **Sampling Strategy**

**Objectives and Reliability Requirements**

The objective is determining the most popular meal in the Innscor chain (Pizza inn, chicken inn, creamy inn) in order to drive more numbers to products that do not have good sales numbers. This improvement will be based on customer needs. The research should take 6 months with sample customer data from several shops spread out across the country.

**Target Population**

The target population is customers who visit the Innscor food chains across the various regions as segmented by the business, for example Mt. Kenya, Nairobi East etc. Only one shop will be considered per region. The population data will not include the Innscor staff that visit the shops. All the customer features such as age, gender will be included in the dataset. The customer names will be omitted so as to ensure customer privacy.

**Sampling Methods**

The sampling method to be used will be stratified sampling. This is a sampling technique where the population is divided into smaller distinct groups that do not overlap but represent the entire population together. While sampling, these groups can be organized and then draw a sample from each group separately.

The customer data will be gathered in groups divided by region. These regions will form our strata groups. From there, one shop will be chosen at random from the different regions.

**Sample Size**

The number of samples will be equal to the total number of regions. These regions are the segmentations done by the business to divide the country for easier divisions and reporting. The parameter of interest is the meal that was purchased. Other observations we expect from the data are age, time\_to\_be\_served, repeat visit, depending on the customer data stored by the food chain. The variance expected will be based on the amount of data. I expect some regions to have more volumes of data than others due to the different population sizes of the different regions.

**Sampling Frame**

The sampling frame is each of the regions segmented by the business needs for reporting.

**2. Data**

**Field Measurements**

N.B this is an assumption of the type of data collected at the different chains

The variables will be:

1. Meal purchased
2. Time to be served
3. Customer mobile number
4. Date
5. Repeat visit
6. Method of payment
7. Store visited e.g. Chicken Inn, Pizza Inn, Creamy Inn

**Quality Assurance / Quality Control**

The Innscor team from the historical data collected through the service provision will provide the dataset. The quality of data provided will be determined by the information provided.

**Analysis**

The data collected will be used to discover the meal preferences of most customers. From this analysis, the chain management will be able to discover which meals to market more to customers and what the customers like.

**3. Implementation**

**Implementation Plan**

The data will be shared from the Innscor databases through the database management teams of the various chains. The management will share the data without compromising the client privacy.

Machine learning experts will do data collection and analysis. The data availed will undergo cleaning and filtering in order to work with data that provides useful sentiments.

The schedule for implementing the project will be 6 months.

|  |  |
| --- | --- |
| Phase | Timelines |
| Client meetup and introduction | 1 month |
| Data sharing and cleaning | 2 months |
| Analysis and recommendations | 3 months |