**Final Presentation Introduction**

The documents are used for the final presentation.

List out three topics summarize the business content field, business questions, and data resources to use.

Each group pick up 1 topic and used the material for reference.

For example, the problem statement and target business questions can be modified or selected based on your group's needs and decisions.

The dataset is large, so please used the filter to select the necessary used columns or parameters, you don’t have to use all the datasets.

|  |  |  |
| --- | --- | --- |
| Health  Biology  Hospitality | **Problem statement:**  The NIH (WHO) used the food death analysis to sponsor the food bank projects in developing and developed countries. The food health analysis helps the organization to use the food and healthy diet to help those countries improve healthy life and reduce the death rate.  Food health and prediction of death rate for data analysis helps the NIH and WHO to optimize the sponsorship and better classify the food supplies and nutrition system of each country, for example, knowing the most demanding and fatal food to provide is crucial to the sponsorship. The analysis used Food\_Supply\_kcal(percentage of energy intake) to help people to improve eating habits to combat spreading diseases.    **Content attributes (parameters)**   * Country * Alcoholic Beverages * Animal fats * Animal Products * Aquatic Products * Cereals * Excluding Beer * Death     **Data Considerations**   * Problem to solve * Data cleaning: any missing values are represented by NA.\? * Data preprocessing: scaling: percentage * Aggregation: sum, count, mean, frequency of death rate across the countries     **Analysis Task**   * Exploratory Data Analysis: * CRISP – DM     **Target Business goals:**   * How accurately the food diet can predict the countries' death rate?   + Regression coefficient * What are the important food diet features that are used to predict the deaths of specific countries? * Ranking 3 most relevant food diet features * Research on why those food diet features are relevant to specific countries but not others? * The average level of Fish, Seafood in developing and developed countries | Food\_Supply\_kcal  And data dictionary |
| Sales, Marketing and Business | **Problem statement:**  Customer data for stock purchasing is a detailed analysis for the financing company to know the purchase behavior of their customers. It helps a business to better understand its customers and makes it easier for them to modify products according to the specific needs, behaviors, and concerns of different types of customers.  Customer data analysis helps a business to modify its product prices based on demanding of customers from different time ranges. For example, instead of reducing the prices to attract the potential customer to the company’s database, a company can analyze which products the customers are most likely to buy and then increase the marketing campaign for the products  **Content attributes (parameters)**   * CustomerID * InvoiceNo * InvoiceDate * Invoice ID * StockCode * Description * Quantity * UnitPrice * StockCode     **Data Considerations**   * Problem to solve * Data cleaning: any missing values are represented by NA.\? * Data preprocessing: price scaling/categorization: low, medium, high (price) * Aggregation: sum, count, mean, frequency of stock price     **Analysis Task**   * Exploratory Data Analysis: * CRISP – DM     **Target Business goals:**   * Which stocks (products) are the most popular? * Who are the purchasers and why they are popular?   o   Research on the most popular 3 stocks, and why   * What are the average spendings for specific products? * What are the variations of the price across the time and how it impacts the purchase rate? | Customer Data:  data\_cus.csv |
| Educational financing and policy | **Problem statement:**  The schools and States want to know about the teachers’ retention rate. The school's financing system used the State reports to obtain information on the expenses of K-12 schools. The school academic system also used the students’ score reports from NAEP data to inform the schools about the student's performance.  School financing and score data analysis help the education system to make the decision on hiring the new teachers and maintain the current teaching force. For example, the high revenue, investment, and expense of the CA state might related to the average scores of the students. The educational system can analyze and consider increasing the teaching hiring and expect a high retention rate in the next school year.  **Content attributes (parameters)**   * State * Enroll * District name * Total revenue * Federal revenue * State revenue * Total expense * Current spending   **Data Considerations**   * Data cleaning: any missing values are represented by NA.\? * Data preprocessing: price scaling/categorization: 1,000 ~ 10,000 * 10,000 ~ 100,000 * 100,000 ~ 1,000,000 * etc * Aggregation: sum, count, mean, frequency of revenue/expense on State levels   **Analysis Task**   * Exploratory Data Analysis: * CRISP – DM   **Target Business goals:**   * Whether or not increase the hiring of the new teachers or increase the education invest the next year? * What are the average levels of the education financing revenue of State A (CA) in comparison to other State (Arizona)? * Are the high revenues related to the students’ average scores? * Tools: stats analysis, plots, tables presentation. * Methods: regression. | NAEP  (National Score  Report card)  data  https://www.kaggle.com/datasets/noriuk/us-educational-finances?select=naep.csv  OR  District.csv |