A green and white logo

Description automatically generated

**CST2213 BUSINESS INTELLIGENCE PROGRAMMING -2 ADVANCED CONCEPTS**

**Project Phase 2: Data Preparation & Initial Analysis**

1. **Data Cleaning and Preprocessing** 
   * **Objective:** Ensure data quality by handling missing values, duplicates, and inconsistencies.
   * **Tasks:**
     + Remove or impute missing values.
     + Normalize and standardize data if necessary.
     + Convert categorical data into numerical formats .
     + Handle outliers using statistical techniques.
   * **Deliverables:**
     + A cleaned dataset ready for analysis.
     + A report on data preprocessing steps taken.
2. **Data Integration and Transformation**
   * **Objective:** Merge multiple datasets and transform them for analysis.
   * **Tasks:**
     + Integrate data from multiple sources into a single dataset.
     + Perform data transformations (e.g., aggregations, feature engineering).
     + Store processed data in a structured format .
   * **Deliverables:**
     + Integrated and transformed dataset.
     + Documentation of transformations applied.
3. **Exploratory Data Analysis (EDA)**
   * **Objective:** Gain insights into data through statistical analysis and visualization.
   * **Tasks:**
     + Generate summary statistics (mean, median, standard deviation, correlations).
     + Visualize distributions using histograms, box plots, and scatter plots.
     + Identify relationships between variables using correlation matrices and heatmaps.
   * **Deliverables:**
     + An EDA report with key findings.
     + Visualizations highlighting important trends and patterns.
4. **Feature Selection and Engineering**
   * **Objective:** Identify and create relevant features for further analysis.
   * **Tasks:**
     + Select key features using statistical tests or machine learning techniques.
     + Create new features based on domain knowledge.
     + Reduce dimensionality if needed.
   * **Deliverables:**
     + A refined feature set for modeling.
     + Justification for selected features.
5. **Preliminary Insights and Hypothesis Testing**
   * **Objective:** Validate initial assumptions and refine the research approach.
   * **Tasks:**
     + Perform hypothesis testing on key relationships.
     + Identify patterns and trends that can be used for predictive modeling.
     + Adjust project scope based on findings.
   * **Deliverables:**
     + A report summarizing initial insights.
     + Recommendations for further analysis in the next phase.

**Submission for Phase 2**

* **Required Deliverables:**
  1. Data cleaning and preprocessing report.
  2. Integrated dataset with transformation documentation.
  3. Exploratory Data Analysis report with visualizations.
  4. Feature selection and engineering documentation.
  5. Summary of preliminary insights and hypothesis testing results.

**Project Phase 2: System Design & Development**

1. **System Architecture Design**
   * **Objective:** Define the system's structure and its components.
   * **Tasks:**
     + Identify system architecture.
     + Define data flow and interactions between system components.
     + Create UML diagrams (Use Case, ERD, Sequence Diagrams).
   * **Deliverables:**
     + System architecture document.
     + UML diagrams outlining key system interactions.
2. **Database Design**
   * **Objective:** Design a structured database schema to store data efficiently.
   * **Tasks:**
     + Define tables, relationships, and constraints.
     + Normalize data to reduce redundancy.
     + Develop ERD (Entity Relationship Diagram).
   * **Deliverables:**
     + Database schema design.
     + SQL scripts for table creation.
3. **Frontend & Backend Development**
   * **Objective:** Implement the core system functionalities.
   * **Tasks:**
     + Develop frontend.
     + Implement backend.
     + Connect frontend and backend.
   * **Deliverables:**
     + Initial frontend and backend implementations.
     + API documentation.
4. **Feature Implementation & Integration**
   * **Objective:** Develop and integrate key system features.
   * **Tasks:**
     + Implement authentication and authorization.
     + Set up data processing logic (e.g., reports, visualizations).
   * **Deliverables:**
     + Functional feature modules.
     + API testing results.
5. **Testing & Debugging**
   * **Objective:** Ensure the system functions correctly before deployment.
   * **Tasks:**
     + Perform **unit testing** (e.g.,PyTest).
     + Conduct **integration testing** to verify module interactions.
     + Identify and fix bugs.
   * **Deliverables:**
     + Test cases and results report.
     + Bug tracking and resolution documentation.

**Submission for Phase 2**

* **Required Deliverables:**
  1. System architecture and database design documents.
  2. Initial frontend and backend implementations.
  3. API documentation.
  4. Feature implementation report.
  5. Testing report and bug fixes.

Siju Philip