GROUP 5

ClearCare Data Initiative

Sponsor: Vijay Koneru

MIS6349.001

Introduction:

The project's purpose revolves around helping people make informed decisions about buying the right insurance according to their needs by highlighting cost variations among insurance providers and hospitals based on their location. The name of the project will be **ClearCare Data Initiative**.

Business Context and Scope Statement:

The healthcare industry presents significant challenges in understanding the costs associated with various insurance providers and healthcare treatment and services. Patients often lack access to transparency in pricing and data on which insurance options would be more economical for specific treatments at different hospitals. The ClearCare Data Initiative aims to bridge this gap by compiling, analyzing, and presenting data that compares the costs of each treatment provided by different insurance providers across hospitals. The end goal is to empower consumers with the information needed to make cost-effective healthcare decisions.

Success criteria:

- **Comprehensive Data Collection**: Successful acquisition of accurate and up-to-date data between multiple hospitals and insurance providers, covering a broad range of treatments and services.
- **Data Accuracy and Reliability**: Ensuring that the presented data is accurate, reliable, and regularly updated to maintain the project's credibility and utility to consumers.
- **Effective Visual Representation**: Development of clear, informative, and interactive visualizations that effectively display cost comparisons and insights for users. These visualizations should enable consumers to easily understand differences in costs between insurance providers for various treatments at different hospitals.

High-level requirements:

1.Data Collection

- **Sources of Data**: Collect healthcare pricing data from hospitals and insurance providers. Focus on publicly available price transparency data, hospital charge lists, and insurance premiums or payment rates.
- **Data Types**: Include data on treatment types, hospital names, geographical location (region, city), insurance provider names, and cost breakdowns (e.g., hospital fees, insurance coverage, out-of-pocket expenses).

2. Data Processing

- **Normalization**: Standardize the data format across different hospitals and insurance providers to ensure consistency in treatment names, cost units, and geographical data.
- **Cleaning and Validation**: Remove duplicates, fill in missing values where possible, and validate the integrity of the data (e.g., checking for extreme outliers or unrealistic cost figures).

3. Data Analysis

- **Cost Comparisons**: Provide comparative analysis across hospitals and insurance providers. Highlight significant cost differences for common procedures or treatments.
- **Insurance Comparison**: Analyze how different insurance plans cover various treatments and the resulting out-of-pocket costs for consumers.

4. Data Visualization

- **Visual Representation**: Create visual representations (e.g., charts, graphs, heatmaps) that illustrate cost variations across hospitals, regions, and insurance providers.
- **User-Friendly Formats**: Ensure that visualizations are easy to interpret, focusing on comparative insights such as lowest cost providers, most expensive regions, etc.

5. Data Documentation

- **Data Access**: Ensure the dataset is available in standard formats (e.g., CSV, Excel) for easy analysis by stakeholders.

6. Stakeholder Feedback

- **Iterative Review**: Collect feedback from key stakeholders (Vijay Koneru) to refine the dataset and visualizations to meet their needs.

List of deliverables:

1.Data Collection Framework

- **Data Sources List**: A comprehensive list of data sources (hospitals, insurance providers, public databases) that will be used to gather pricing information.
- **Data Collection Templates**: Standardized templates for collecting data (e.g., data fields for treatment names, costs, insurance provider details, etc.).

2. Raw Dataset

- **Initial Data Dump**: Raw, unprocessed data collected from all sources, including information on treatment costs, insurance coverage, and hospital details.
- **Data Dictionary**: A document explaining the structure, fields, and meanings of each data point in the raw dataset.

3. Cleaned and Normalized Dataset

- **Processed Dataset**: A cleaned and standardized dataset, ready for analysis. This should include normalized treatment names, consistent cost units, and resolved data gaps.

4. Visual Representations

- **Charts and Graphs**: Visual representations of the data, such as bar charts, heatmaps, and line graphs, illustrating cost variations across hospitals and insurance providers.

5. Final Dataset

- **Public Dataset**: A finalized version of the dataset available in user-friendly formats (e.g., CSV, Excel), cleaned, validated, and documented for ease of use.

6. Project Summary and Presentation

- **Presentation Deck**: A slide deck summarizing the project's objectives, process, key insights, and final deliverables for presentation to stakeholders/interested parties.

These deliverables will ensure the successful completion of the ClearCare Data Initiative, providing valuable insights into healthcare cost transparency.

Milestone Schedule:

All the milestones will be completed using the tech-stack mentioned at the end.

Milestone	Description	Expected Completion
Project Kickoff	Finalize scope, objectives, and stakeholder expectations	Week 1
Data Sources Identified	Comprehensive list of data sources (hospitals, insurance providers, public databases)	Week 2
Data Collection Framework	Develop standardized data collection templates for pricing information (hospitals, insurance, etc.)	Week 3
Raw Data Collection Complete	Collect all raw data from identified sources (initial data dump)	Week 4

Data Dictionary Created	Documentation explaining the structure and fields of the raw dataset	Week 5
Data Cleaning & Normalization	Process data to remove duplicates, normalize terms, handle missing values, and resolve inconsistencies	Week 6
Data Validation Report	Document steps taken during cleaning and validation, including handling outliers	Week 7
Initial Data Analysis	Perform comparative cost analysis across hospitals and insurance providers	Week 8
Data Visualization Draft	Develop visual representations (charts, graphs, heatmaps) to illustrate cost variations	Week 9
Stakeholder Feedback	Collect feedback from stakeholders (Vijay Koneru) to refine dataset and visualizations	Week 10
Final Dataset	Create and finalize the public dataset, including cleaned data, documentation, and user-friendly formats	Week 11
Visual Representations Final	Finalize visualizations based on stakeholder feedback and final data	Week 12
Project Presentation	Develop and present a project summary deck to stakeholders	Week 13

Assumptions and Constraints:

Assumptions:

- **Availability of Data**: It is assumed that hospitals and insurance providers will make the necessary pricing data available through public price transparency policies or accessible databases.
- **Consistency of Data**: It is assumed that the pricing data collected from hospitals and insurance providers will be structured in a way that allows for standardization and comparison (e.g., uniform treatment definitions, cost units, etc.).
- **Legal and Ethical Compliance**: It is assumed that all data collected and used complies with legal requirements, including HIPAA regulations and data privacy laws.

Constraints:

- **Data Quality and Accuracy**: There may be discrepancies, inconsistencies, or errors in the data provided by different sources, affecting the accuracy of the final data set.

- **Time and Resource Limitations**: The project may face time or budget constraints, limiting the depth and breadth of data collection, analysis, or visualization.
- **Geographical Scope**: The scope of the project may be limited to certain regions or states due to varying regulations on price transparency across different regions.

Risks:

- **Misinterpretation of Data**: End users (patients or policymakers) may misinterpret the data presented, leading to uninformed or inappropriate healthcare decisions.

- **Inconsistent Data Formats**: If hospitals and insurance providers report costs in inconsistent formats (e.g., different units, treatment names, or definitions), it may create challenges in normalizing and comparing the data.

Team Structure:

The ClearCare Data Initiative uses a flat team structure to keep things simple and encourage teamwork. Everyone on the team is responsible for their part and works together without strict roles or hierarchy. This helps us communicate faster, make decisions quickly, and solve problems more easily. Each team member takes ownership of their tasks, and we all share responsibility for the project's success.

Communication plan:

- Weekly progress meetings with the team and mentor
- Communication tools: Teams for project progress discussions and outlook for emails and communication

Approvals:

Stakeholders: Team 5 (Application Development Team), Vijay Koneru (Project Mentor)

Techstack:

- Data Cleaning, Data Loading, Normalisation: Python
- Database: PostgreSQL
- IDE: Jupyter/Google collab