

**MediPriceInsight – Project Charter**

**BUAN 6390.001 - Analytics Practicum**

**Professor**

Mandar Samant

**Project Mentor**

Vijay Koneru

**Group 3 Members**

Tarun Raghu - TXR230002

Pratham Khubchandani - PXK230013

Krishna Apurva - KXA230007

Apoorva Shukla - AXS220598

Shanvi Shivali Kothapalli - SXK220320

Siva Srinivas Narra - SXN230069

Rahul Sadineni - RXS230051

Aarthik Kothuri - AXK230022

Naga Harshath Medisetti - NXM220059

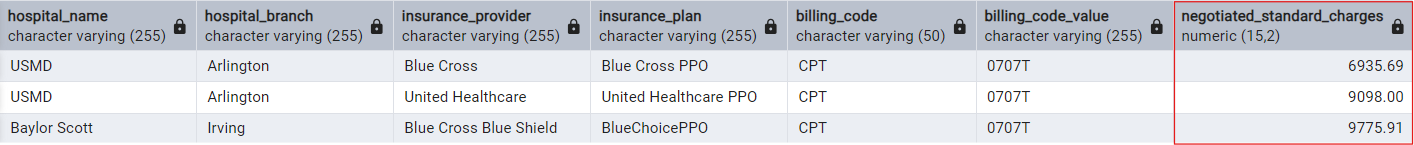
**PROJECT CHARTER**

**1. INTRODUCTION**

Understanding the cost of medical care can be overwhelming due to the lack of transparent and easily accessible pricing information for treatments from hospitals and the complexities introduced by insurance providers. MediPriceInsight aims to address this issue by collecting, analyzing, and presenting publicly available pricing data for treatments. By considering factors such as location, insurance provider, insurance plan, hospital branch and treatment cost, our platform empowers individuals with the knowledge they need to make cost effective healthcare decisions.

**2. a. BUSINESS CONTEXT AND SCOPE STATEMENT**

The U.S. healthcare system is notoriously complex, with significant variations in costs across hospitals, insurance providers, and geographical locations. Patients often struggle to make informed financial decisions about their medical care due to a lack of pricing transparency. For instance, consider a bone repair injection for joint defects at USMD Arlington and Baylor Scott at Irving, which are approximately 20 miles apart (a 22-minute drive). (Appendix 1) Even when considering only PPO plans for this specific procedure, there is substantial variation (41%) in negotiated standard charges under different insurance plans.

The final deliverable will be interactive visualizations to help users compare costs across providers and hospitals, along with a structured dataset for analysis enabling them to make informed decisions based on data-driven insights.

**2. b. MARKET RESEARCH**

**i. Hospital price transparency**

A CMS initiative aimed at empowering patients with the information necessary to make informed healthcare decisions. Effective January 1, 2021, hospitals in the United States are mandated to publicly disclose their standard charges for all items and services they provide. (Appendix 2)

**ii. Market Demand for Healthcare Cost Transparency**

According to a 2022 survey by Gallup and West Health (Appendix 3), 38% of Americans reported skipping medical care due to high costs, highlighting the urgent need for transparent healthcare pricing information. Several platforms, such as GoodRx and FAIR Health exist but with limitations.

**iii. Competitive Landscape**

GoodRx focuses on drug pricing, while FAIR Health provides cost estimates with limited hospital data. However, none of these platforms deliver a fully integrated solution that combines comprehensive hospital pricing with detailed insurance provider breakdowns. This makes MediPriceInsight a unique and valuable addition to the market.

**3. SUCCESS CRITERIA**

The success of our solution depends on the following factors:

**i. Comprehensive Data Collection**

Covering a larger number of hospitals in the assigned states will enhance our success. Data will be sourced from at least 10 healthcare providers in each major city assigned, ensuring that every area or pin code is satisfied.

**ii. Accuracy and Reliability**

Minimize errors and prevent data loss during file ingestion. Perform imputation to ensure completeness and usability for the end user. Present the data without manipulating existing costs, ensuring all data is verifiable.

**iii. Technical Excellence and Scalability**

Demonstrate proficiency with the latest open-source technologies for managing big data. Incorporate error handling and use interactive visual tools to present clear cost comparisons and insights. Additionally, leverage AI to speed up development and handle various scenarios of ingestion depending on the file structure. Develop a scalable solution that can be extended to other states.

**iv. Documentation**

Ensure the project is well-organized, thoroughly documented, easily interpretable, and reproducible. Provide all the information needed to understand the work done and to effectively use the MediPriceInsight tool.

**4. HIGH LEVEL REQUIREMENTS**

Develop a comprehensive application with a Price Transparency and Cost Comparison Tool to visualize hospital procedure pricing and insurance plan costs based on user location. It will also feature a Data Ingestion Tool with a user-friendly UI for inputting machine-readable files, ensuring scalability and seamless database integration. The platform-independent bundle will include all necessary source files and a consolidated database, along with a detailed user manual for setup, usage, and troubleshooting.

**5. LIST OF DELIVERABLES**

**i. Price Transparency and Cost Comparison Tool**

Visualize hospital procedure pricing and insurance plan costs based on user location. Compare various standard for each insurance provider and plan to help users make cost-effective decisions.

**ii. Data Ingestion Tool**

Develop a user-friendly, form-based UI for inputting machine-readable files and ingestion strategies. Ensure scalability to add more files to the database without location restrictions.

**iii. Application**

Bundle the Data Ingestion Tool and Cost Comparison Tool into a platform-independent application. Include raw source files and a consolidated database, ensuring it can be set up and run on any hardware.

**iv. User Manual and Live Demonstration of Product**

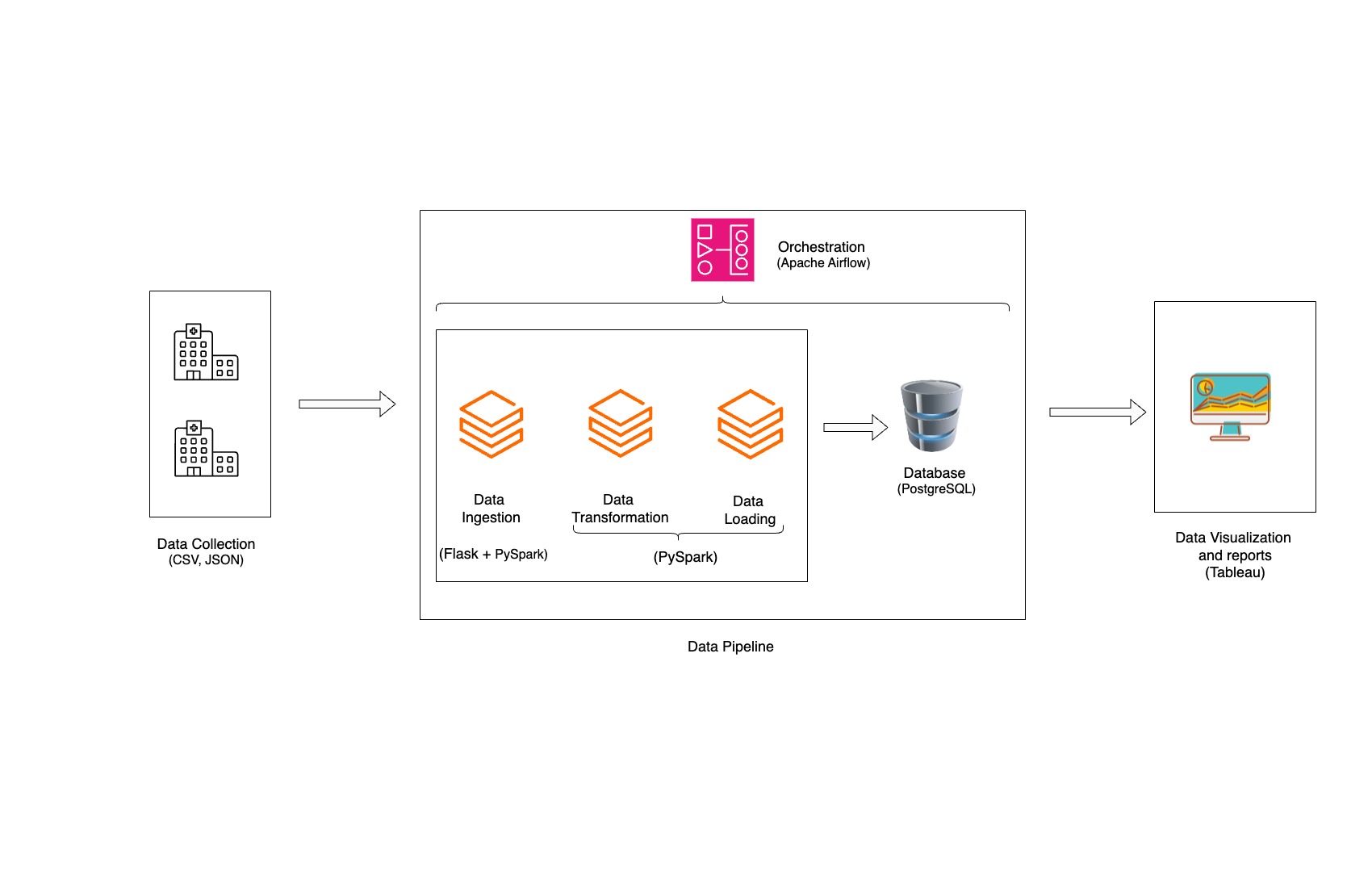
A comprehensive user manual will guide users through setup, usage, and FAQs. The live demonstration will showcase key features of the application, followed by an interactive Q&A session.

**6. TECHNICAL DETAILS**

**a. Approach**

The project involves collecting raw hospital price transparency data from machine-readable files (JSON, CSV) available on hospital websites and storing them locally. A form-based UI is used to input and submit information about the files and ingestion strategy, which is communicated to Python scripts via Flask. PySpark is employed to execute ETL processes, ensuring the data is cleaned, normalized, transformed, consolidated, and validated before being stored in a PostgreSQL relational database. Power BI is used to visualize and generate reports from the processed data. The entire workflow, including data ingestion, transformation, file validation, and database updates, is automated using Airflow. Finally, the application is containerized with Docker for seamless deployment.

**b. High Level Architecture and Design**



**c. Technology and its Purpose**

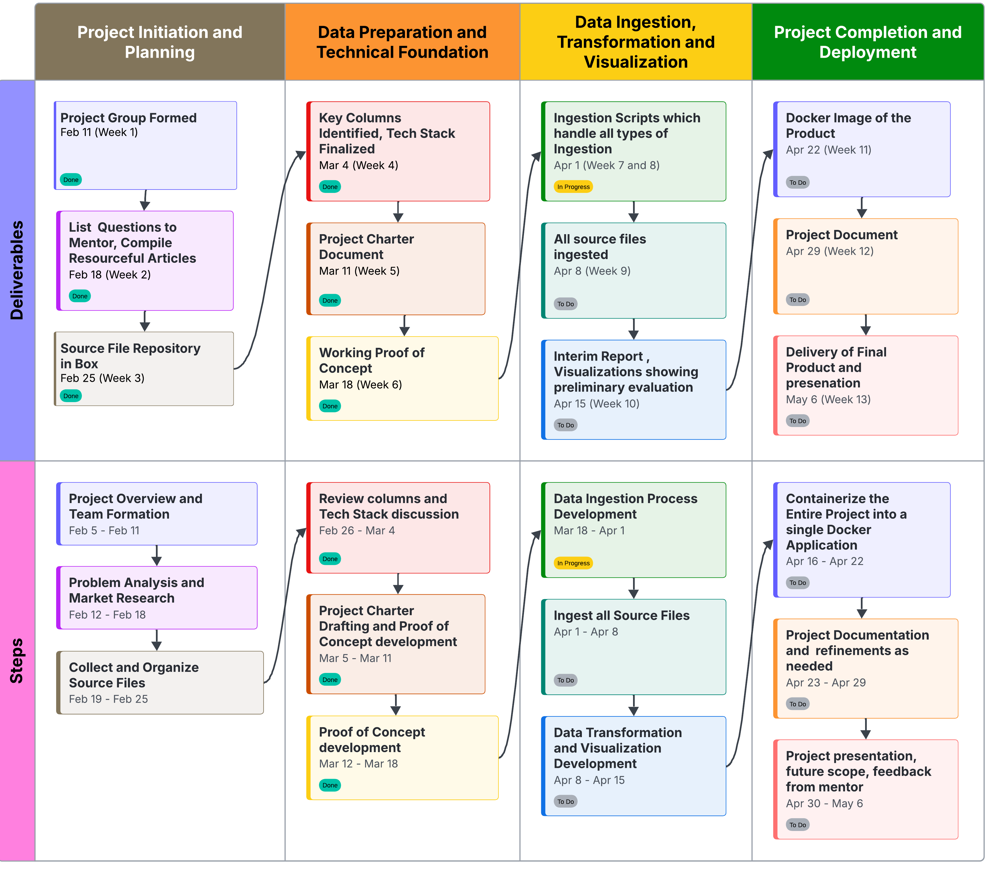
|  |  |  |
| --- | --- | --- |
| Technology | Purpose | |
| PySpark | | Data Transformation, cleaning, validating, importing data from flat file |
| PostgreSQL | | Stored raw and processed data, ensuring indexing and constraints |
| HTML | | To create a form-based UI for entering details, displaying reports |
| Flask | | To help with front-end back-end communication |
| Apache Airflow | | Orchestrate the workflow |
| Power BI | | Data visualization, interactive report dashboards for cost comparison. |
| Docker | | Containerization of data pipeline scripts for consistency across environments. |
| Cursor AI | | For building HTML, Python Scripts |

**d. Technical Limitations**

i. The data is static and not real-time, which means there is a possibility of stale information being present in the reports.

ii. Some files are large and may take considerable time to process.

**7. MILESTONE SCHEDULE**



**8. ASSUMPTIONS AND CONSTRAINTS**

1. Healthcare cost and insurance data have been collected from public sources is assumed to be accurate.
2. Only selected hospitals in major cities (Anaheim, DFW, Long Beach and Arlington) will be considered.
3. Only CPT codes will be used for treatment identification to ensure standardization across hospitals.
4. Web scraping is not used, data is manually collected from hospital websites.
5. The system will not automatically detect or track changes in pricing data.
6. The data will be static and not real time.
7. Variables such as co-pays, deductibles, and out-of-pocket costs are not in scope for the project.

**9. RISKS**

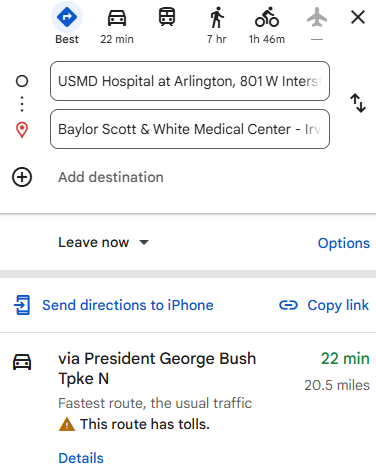
1. **Non-compliance with Transparency Laws**: Many hospitals do not fully comply with price transparency regulations, leading to incomplete or inaccurate data.
2. **Inaccurate or Outdated Data**: Hospitals and insurance providers may not update their datasets regularly, resulting in outdated cost data.
3. **Inconsistent Data Formats and Reporting Standards**: Different formats, coding systems, and reporting standards across states make data integration and analysis challenging.
4. **Consumer Resistance**: Skepticism about the accuracy and utility of the information, especially in regions with historically opaque healthcare pricing, can lead to low adoption rates and consumer hesitance to trust the data.

**10. TEAM STRUCTURE**

|  |  |  |  |
| --- | --- | --- | --- |
| Sno | Name of Resource | Role | Responsibilities |
| i | Vijay | Mentor | Guide the team members to success |
| ii | Tarun | Project Manager | Oversee project planning and execution |
| iii | Krishna | ETL Developer | Design and Implement ETL processes |
| iv | Pratham | Reporting Analyst | Develop Reports, Gather Insights |
| v | Apoorva | ETL Developer | Design and Implement ETL processes |
| vi | Shanvi | Reporting Analyst | Develop Reports, Gather Insights |
| vii | Srinivas | Data Analyst & DevOps Engineer | Deploy application and gather insights |
| viii | Aarthik | Data Collection & UI development | Collect data and develop user interfaces |
| ix | Rahul | Quality Analyst | Ensure data quality and completeness |
| x | Harshath | Database Developer | Design and maintain database structure |

APPENDIX

1. Travel distance and time from USMD at Arlington and Baylor Scott at Irving using Google Maps.



2.[**https://www.cms.gov/priorities/key-initiatives/hospital-price-transparency/hospitals**](https://www.cms.gov/priorities/key-initiatives/hospital-price-transparency/hospitals)

3.<https://westhealth.org/news/new-study-reveals-more-struggling-to-afford-healthcare/>