Deployment Plan Document

# 1. Introduction

Project Overview: The goal is to create a web application that provides insights on medical provider ratings across the US. This project will help patients, healthcare providers, insurers, and administrators make informed decisions based on comprehensive data.

Team Members: Nabeel Bahrawi, Jeffery Liu, Jacky Jiang

# 2. Objectives

Provide insights/services to improve client experiences.

Allow users to make data-driven decisions.

Continuously refine the tool with new data and feedback.

# 3. Scope

Data Acquisition: Source data from RateMDs.

Data Transformation: Clean and transform raw data into actionable data.

Data Warehousing: Structure data using a star schema into data warehouse.

Visualization: Develop a user interface to visualize the data and generate insights.

# 4. Deployment Architecture

Data Source: RateMDs (2.2 million records, ~2.39GB)

Storage: MongoDB for raw and transformed data storage

Data Warehouse: Implement star schema with five dimensions (doctor, rating, location, location city, location rating) and a fact table.

Data Warehouse Tool: Snowflake

ETL Pipeline: Extract, Transform, Load process to move data from MongoDB to Snowflake.

Visualization Tool: Domo

# 5. Deployment Steps

## Step 1: Environment Setup

Development Environment: Set up local development environment with necessary tools (Python, Jupyter Notebook, MongoDB, etc.).

Production Environment: Set up cloud infrastructure (AWS, Azure, or Google Cloud) with virtual machines, databases, and necessary services.

## Step 2: Data Acquisition

Extract data from RateMDs and load it into MongoDB.

Ensure data integrity and handle any duplicates or missing values.

## Step 3: Data Transformation

Clean and transform the data using Jupyter Notebook or Alteryx.

Implement necessary data profiling and cleaning steps.

## Step 4: Data Warehouse Setup

Design and implement the star schema in the data warehouse.

Load the transformed data into the warehouse.

## Step 5: ETL Pipeline

Develop ETL scripts to automate the data extraction, transformation, and loading process.

Schedule regular data updates to keep the data warehouse current.

## Step 6: Visualization

Create dashboards and reports in Domo to visualize the data.

Implement filters and interactive elements for user exploration.

## Step 7: Deployment

Deploy the web application and dashboards on the cloud.

Ensure the application is accessible and perform thorough testing.

## Step 8: Monitoring and Maintenance

Set up monitoring tools to track the performance and usage of the application.

Implement a feedback loop for continuous improvement based on user input.

# 6. Risk Management

Data Security: Ensure data is securely stored and accessed, following best practices and compliance requirements.

Performance: Optimize ETL processes and database queries to handle large volumes of data efficiently.

Scalability: Design the system to scale with increasing data and user load.

# 7. Timeline

Month 1: Environment setup, data acquisition, and initial data transformation.

Month 2: Data warehouse setup and ETL pipeline development.

Month 3: Visualization development and initial deployment.

Month 4: Testing, optimization, and final deployment.

# 8. Conclusion and Next Steps

Continue integrating new data sources and optimizing the performance of the data warehouse.

Expand the application's features based on user feedback and emerging needs.