**IST 722: Data Warehouse: Project Charter**

**Project: Creating a Data Warehouse, ETL Pipeline and Reporting Tool for the New York Metro Authority**

**Team: The Special Ones**

**Members:**

Keerthi Krishna Aiyappan, Sai Swetha Lakkoju, Venkata Sai Namana, Harika Gangu

**Business Processes:**

**1. Rider Volume Tracking**

* **Improved Resource Allocation**: Helps identify high traffic stops and lines, enabling better deployment of resources such as additional trains during peak hours.
* **Revenue Optimization**: Tracks fare revenue associated with rider volumes and helps forecast future earnings.
* **Service Planning**: Provides insights into rider demand patterns, informing decisions on expanding routes or adjusting schedules.

**2. Stop Utilization Analysis**

* **Infrastructure Efficiency**: Measures how effectively each stop is being utilized, helping prioritize maintenance or upgrades for heavily used stops.
* **Operational Cost Management**: Reduces costs by identifying underutilized stops that may need schedule adjustments or alternative transport modes.
* **Customer Experience**: Ensures stops with high usage are equipped with adequate amenities like seating, shelters, and lighting.

**3. Line Performance**

* **Service Reliability**: Tracks delays, wait times, and on-time performance to improve schedule adherence and reliability.
* **Capacity Management:** Evaluates whether the number of vehicles and their capacity on each line meets passenger demand.
* **Public Satisfaction:** Identifies problem areas in service delivery, leading to better customer satisfaction by addressing bottlenecks.

**4. Remote Unit Monitoring**

* **Proactive Maintenance:** Monitors the health of remote devices like ticketing machines or digital displays, reducing downtime.
* **Operational Uptime**: Ensures critical equipment like signal systems and sensors are functioning, improving safety and efficiency.
* **Cost Savings:** Prevents costly failures through predictive maintenance by tracking health metrics like battery levels.

**5. Rider Segmentation**

* **Targeted Marketing**: Segments riders based on demographics or travel patterns, enabling personalized offers or campaigns.
* **Enhanced Service Design**: Tailors services to the needs of specific rider groups, like students, seniors, or commuters.
* **Policy and Planning**: Provides data for government or operator policies that focus on inclusivity and accessibility for diverse rider groups.

**Project Timeline:**

**Weeks 1 and 2: Requirements and Planning**

* Data Profiling
* Business Process identification
* Bus Matrix Creation

**Weeks 3 and 4: Initial Design**

* Dimensional Modelling
* Source to Target Mapping
* Technical Architecture Design

**Weeks 5, 6, 7: Implementation**

* Snowflake Setup
* Schema Creation
* ETL Development with dbT

**Weeks 8,9,10: Analytics and Testing:**

* Power BI Dashboard Development
* Testing and Validation
* Documentation

**Weeks 11 and 12:**

* Video presentation creation
* Final Documentation
* Project Delivery

**Weeks 13 and 14: (Optional)** : To have extra time just in case.

**Technical Stack:**

* **Data Warehouse**: Snowflake
* **ETL Tool:** dbT
* **Orchestration:** Airflow (if needed)
* **BI Tool:** Power BI
* **Version Control:** Git

**Deliverables:**

1. Project Documentation
2. Bus Matrix
3. Dimensional Models
4. Source to Target Mappings
5. ETL Code
6. BI Dashboards
7. 10-minute presentation video

**Primary Roles: (To be Filled)**

* Harika Gangu:
* Shweta Lakkoju:
* Venkata Sai Namana:
* Keerthi Krishna Aiyappan:

**FUNCTIONAL REQUIREMENTS:**

|  |  |  |
| --- | --- | --- |
| **Category** | **Requirement** | **Type** |
| **Data Collection** | Collect real-time data on rider entries, exits, line delays, and stop utilizations. | Functional |
|  | Integrate with sensors, ticketing systems, GPS, and remote monitoring devices. | Functional |
|  | Support historical data import for trend analysis and forecasting. | Functional |
| **Data Processing** | Process data from multiple sources and ensure its integrity. | Functional |
|  | Store processed data in a relational database with efficient indexing for queries. | Functional |
| **Reporting** | Provide real-time dashboards to monitor key metrics. | Functional |
|  | Generate periodic reports for stakeholders. | Functional |
| |  | | --- | | **Alert System** | | |  | | --- | | Send notifications for high rider volumes, device malfunctions, or delays. | | Functional |
|  | Generate predictive alerts for maintenance requirements. | Functional |
| **Segmentation** | Segment riders by demographic, travel patterns, and time-of-day behavior. | Functional |
| **Service Optimization** | |  | | --- | | Identify underutilized stops and recommend schedule adjustments. | | |  |  | | --- | --- | |  | Functional | |
|  | |  |  | | --- | --- | |  | Detect overcrowded stops and recommend adding capacity or rerouting services. | | |  |  |  | | --- | --- | --- | |  |  | Functional | |
| |  |  | | --- | --- | | **Integration** |  | | |  |  |  | | --- | --- | --- | |  | Integrate seamlessly with new data sources, third-party systems and APIs. |  | | |  |  | | --- | --- | |  | Functional | |

**NON FUNCTIONAL REQUIREMENTS:**

|  |  |  |
| --- | --- | --- |
| **Category** | **Requirement** | **Type** |
| **Scalability** | Handle large volumes of data across stops, riders, and remote units. | Non-Functional |
|  | |  | | --- | |  |  |  | | --- | | Support scaling to accommodate new cities or regions. | | Non-Functional |
| **Performance** | Ensure real-time data processing with minimal latency. | Non-Functional |
|  | Perform analytics on large datasets within seconds. | Non-Functional |
| **Reliability** | Implement fault-tolerant mechanisms to handle failures. | Non-Functional |
|  | Guarantee high uptime (e.g., 99.9%) for critical systems. | Non-Functional |
| **Data Security** | Encrypt data in transit and at rest to protect rider and operational data. | Non-Functional |
|  | |  | | --- | |  |  |  | | --- | | Implement role-based access control (RBAC) for authorized personnel. | | Non-Functional |
| **Data Acuracy** | Validate incoming data to ensure consistency and accuracy. | Non-Functional |
|  | |  | | --- | |  |  |  | | --- | | Regularly audit data pipelines to prevent anomalies. | | Non-Functional |
| **Usability** | Ensure dashboards and interfaces are user-friendly and intuitive. | Non-Functional |
|  | Provide customizable filters and views for different user roles. | Non-Functional |
| **Compliance** | Adhere to data protection laws like GDPR or CCPA. | Non-Functional |
|  | Meet accessibility standards for government or public agency use. | Non-Functional |
| **Interoperability** | Support data exports in standard formats (e.g., CSV, JSON). | Non-Functional |
|  | Allow integration with existing transportation systems or tools. | Non-Functional |
| **Maintainability** | Ensure modular architecture to simplify updates or additions of features. | Non-Functional |
|  | Provide clear documentation for system administrators and developers. | Non-Functional |
| **System Availability** | Provide 24/7 availability with robust failover mechanisms. | Non-Functional |
|  | Schedule maintenance during non-peak hours with minimal disruption. | Non-Functional |