**End-to-End Data Integration and Visualization with Snowflake and Tableau**

**Overview of the project:**

The goal of the project is to import raw data into Snowflake and create a staging schema to do further analysis in data and export the results to the production schema in Snowflake. Next, I connected production schema in Snowflake to Tableau to create interactive dashboards and visualizations and then I published in my Tableau publish for wider sharing and access.

**Goals and Objectives:**

The main objectives of this project were:

Importing of Data: Importing large datasets efficiently into Snowflake and ensure the loaded was ready for the further analysis

Staging Schema Creation: Creating the staging schema with 3-dimension tables along with the fact table where the raw data cleansed , transformed and prepared for further analysis before being moved to the production schema.

Data Analysis and Export: Perform data analysis on the staging schema, then export the cleaned and transformed results into the production schema.

Tableau Integration: To connect Snowflake's production schema to Tableau for data visualization.

Publish Dashboards: Finally, create and publish interactive dashboards in Tableau Public for sharing insights.

**Process and Execution :**

I created a database warehouse, database, and created staging schema. I created a 3-dimensional table and one fact table. Next the task involved loading data from multiple sources into Snowflake which matches to staging tables. This included ensuring the data was structured properly and without inconsistencies before beginning analysis. Once the data is loaded staging schema allowed us to perform necessary data cleaning, transformation, and validation processes without affecting the production schema. After data cleansing and transformation in the staging schema, I created 3 dimensional tables and one fact tables . Then I moved the results into the production schema for final reporting and analysis. Next, I connected the production schema in Snowflake to Tableau by configuring the data source connection. This integration allowed Tableau to pull data directly from Snowflake. Using Tableau, I created interactive visualizations to represent key insights and metrics. I incorporated various visualization techniques such as bar charts, line graphs, and maps for a well-rounded presentation. After finalizing the dashboards, I published them to Tableau Public, ensuring they were accessible to stakeholders and the broader community for feedback and decision-making.

**Challenges I faced during this Project:**

**Data Cleaning and Analysis:**

I faced a few challenges in data cleaning and transformation in staging schema of the snowflake includes handling missing values and formatting inconsistencies, which impacted data integrity. Data type mismatches caused disruptions in transformations. These challenges emphasized the importance of rigorous data validation and standardized preprocessing for reliable analysis.

**Issues in Connection:**

I faced a few challenges in connectivity issues between Snowflake and Tableau, particularly with data refreshments. This required adjustments in the Snowflake configuration and Tableau data connection settings.

**Outcomes of the Project**

The project was a success in terms of the integration between Snowflake and Tableau. The staging schema setup allowed for efficient data processing, ensuring clean and accurate data in the production schema. The dashboards created in Tableau were well-received by stakeholders, who found the visualizations insightful and easy to interpret. Tableau Public allowed for wider dissemination of the insights, helping external stakeholders and the community gain access to key business metrics. The project was completed within the agreed timeframe, and the dashboard was successfully deployed for ongoing use.

**Lessons Learned from the Project**

One of the key takeaways was the importance of thorough data cleansing and transformation in the staging schema. This step is essential to ensure the accuracy of any analysis that follows. I learned that both Snowflake and Tableau require regular performance optimizations. For Snowflake, tuning queries were essential. For Tableau, ensuring data sources were well-aggregated and dashboards were not overloaded with real-time data was critical for performance. While the integration between Snowflake and Tableau worked well overall, ensuring a stable connection between the two platforms is important for consistent data updates and visualization Creating insightful, interactive dashboards was key to ensuring stakeholders could easily interpret and explore data. The ability to filter the data was highly valued by the users.

**Areas for Improvement in future:**

In the future, I will explore ways to automate the data transformation process in the staging schema to reduce manual intervention and improve efficiency. While the dashboards were effective, I could have made better use of Tableau’s advanced features such as dynamic dashboards or storylines to present the data in a more engaging manner. I would implement more robust error handling for the connection between Snowflake and Tableau to avoid data refresh failures and ensure continuous data synchronization. Apart from this I try to implement the same project in cloud environment like Azure to get knowledge about that environment.

**Dashboard Link:**

[**https://public.tableau.com/views/DataEngineringProject/EndtoEndOrderAnalyticsDashboard?:language=enUS&publish=yes&:sid=&:redirect=auth&:display\_count=n&:origin=viz\_share\_link**](https://public.tableau.com/views/DataEngineringProject/EndtoEndOrderAnalyticsDashboard?:language=enUS&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link)

**Conclusion:**

This project provided valuable experience in data integration, analysis, and visualization using Snowflake and Tableau. The creation of a staging schema facilitated clean and accurate data processing, while the interactive Tableau dashboards offered stakeholders a powerful tool for decision-making. Although there were some challenges in data cleansing and system performance, these were successfully addressed, and the project was delivered on time. I look forward to applying these lessons in future projects to further refine the process and enhance my skills in data integration and visualization.