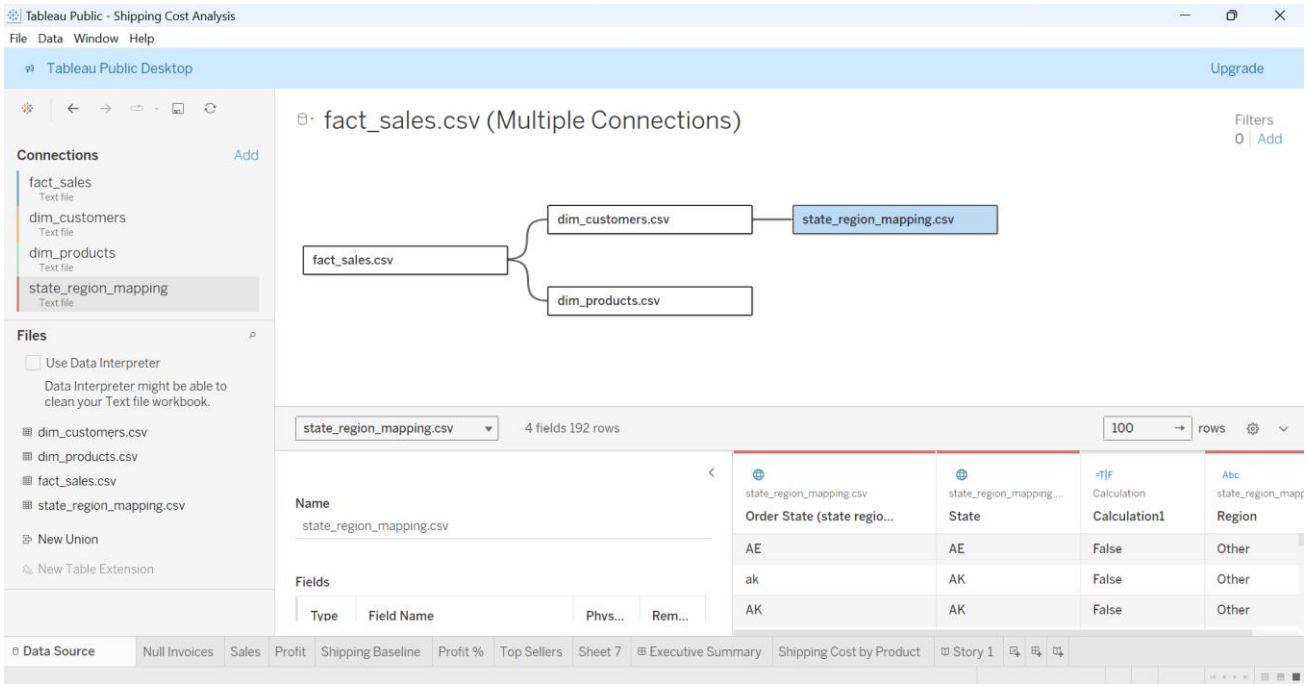

Portfolio Documentation: Shipping Cost Analysis Data Model

Project Title: Shipping Cost Analysis Data Integration in Tableau

1. Overview

This project demonstrates a comprehensive data integration and modelling process in Tableau Public to support a shipping cost analysis. It involves joining multiple data sources, cleaning and transforming data, and creating a robust relational data model for visualization and insights.

2. Data Model Structure

The data model is built using multiple connections between fact and dimension tables to create a star schema-like structure. This model enables seamless data exploration and analysis of shipping costs across different dimensions such as products, customers, and regions.

3. Data Sources:

a. fact_sales.csv:
   - Contains transactional sales data.
   - Includes fields such as Order ID, Shipping Cost, and Product ID.

b. dim_customers.csv:
   - Dimension table with customer attributes.
   - Key fields: Customer ID, Order State, Postal Code, and geographic details like Latitude and Longitude.

c. dim_products.csv:
   - Dimension table with product information.
   - Key fields: Product ID, Description, Stock Code, and Weight.

d. state_region_mapping.csv:
   - Mapping table to associate states with regions.
   - Key fields: State, Region.

4. Data Integration Process

The data integration is accomplished through logical relationships defined in Tableau Public:

a. Join Logic:

- fact_sales.csv is the central fact table.
- dim_customers.csv is joined to fact_sales.csv using the Customer ID field.
- dim_products.csv is joined to fact_sales.csv using the Product ID field.
- state_region_mapping.csv is joined to dim_customers.csv using the Order State field to map states to regions.

b. Joins and Relationships:

- Joins are set up in Tableau to ensure referential integrity between the fact and dimension tables, enabling robust drill-down analysis.
- Example join conditions:
  - fact_sales.CustomerID = dim_customers.CustomerID
  - dim_customers.OrderState = state_region_mapping.State

5. Data Cleaning

a. State Standardization:

- The Order State field in state_region_mapping.csv contains mixed-case values (e.g., "ak", "AK").
- Cleaning is performed to standardize state names for accurate joining.

b. Field Transformation:

- Calculated fields were created for data validation and transformation:
  - Example Calculation: Flagging inconsistencies in state mapping.

```
IF [Order State] IS NULL THEN "Unknown"
ELSE "Valid"
END
```

c. Field Renaming:

- **Renamed fields for clarity (e.g., state_region_mapping.State → State).**

6. Key Features of the Data Model

a. Scalability:

- The model supports additional dimensions, such as warehouse locations or product categories, without major restructuring.

b. Dynamic Mapping:

- The state_region_mapping.csv table allows dynamic regional aggregation for analyses such as shipping cost by region.

c. Normalization:

- Splitting customer, product, and region data into dimension tables ensures a clean and efficient model.

7. Purpose and Applications

The integrated data model is designed to support the following analytical objectives:

a. Compare baseline and what-if shipping costs by product, customer region, and state.

b. Drill down into shipping cost trends by region or product type.

c. Identify regions or products contributing most to shipping costs.

d. Perform scenario analyses for logistics optimization.

8. Challenges and Solutions

a. Inconsistent State Names:

- Solution: Standardized state names using calculated fields and verified mappings in Tableau.

b. Complex Joins:

- Solution: Visualized and debugged relationships using Tableau's Data Model Canvas to ensure correct joins.

c. Performance:

- Solution: Limited data fields and applied filters to reduce data load during dashboard visualization.