

Name: Hussien Mohamed Ahmed

Data Warehouse Project

1. Introduction

The project involves designing and implementing a **Data Warehouse (DWH)** for the **Gravity Books** database.

The goal was to:

- Extract, transform, and load (ETL) data from the OLTP database into a structured DWH.
 - Apply **dimensional modeling techniques** (snowflake schema).
 - Handle **Slowly Changing Dimensions (SCD)** for historical tracking.
 - Build a **Power BI dashboard** for business insights.
-

2. Source Database (OLTP)

The source system is the gravity books database, which represents the online transactional processing (OLTP) environment.

It is normalized and optimized for day-to-day transactions, not for analytics.

List of Tables in OLTP

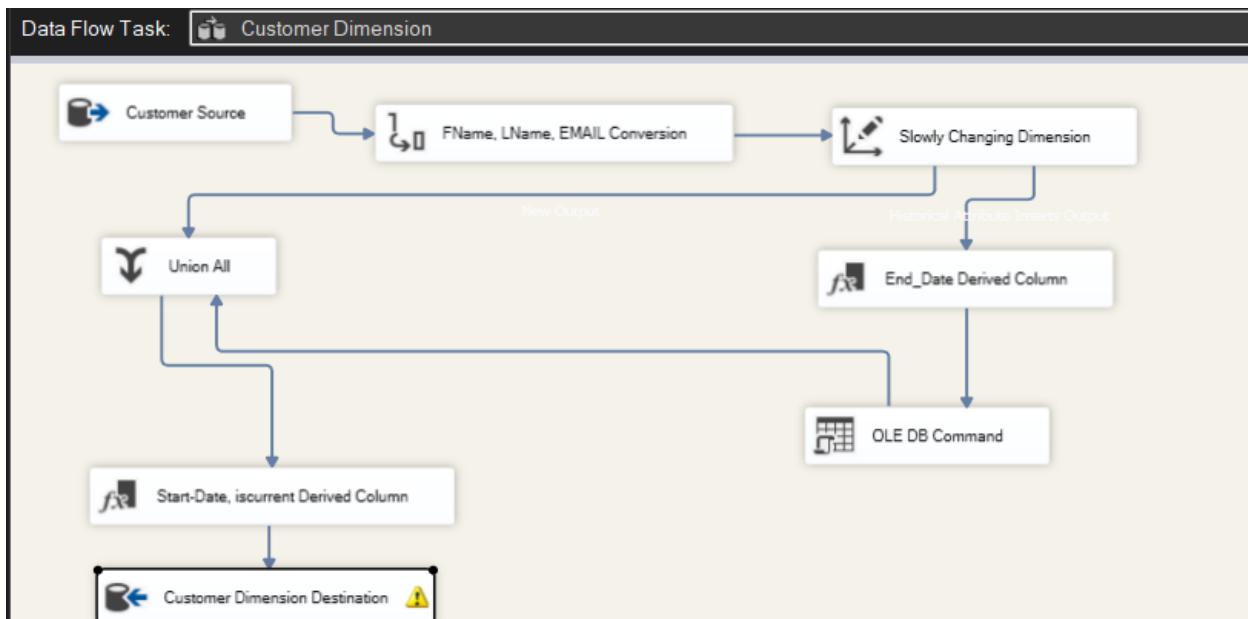
1. **address** → stores customer addresses (street, city, country, etc.).
2. **address_status** → maintains status information for addresses.
3. **author** → details of book authors.
4. **book** → book details (title, ISBN, publisher, language, publication date, etc.).
5. **book_author** → bridge between authors and books (many-to-many relationship).

- 6. book_language → defines available book languages.**
 - 7. country → list of countries for addresses and customers.**
 - 8. cust_order → main order table (customer purchases).**
 - 9. customer → information about customers (names, emails, etc.).**
 - 10. customer_address → bridge linking customers and their addresses.**
 - 11. order_history → historical log of orders (status changes, etc.).**
 - 12. order_line → details of each item inside an order (books purchased per order).**
 - 13. order_status → defines different statuses for orders (pending, shipped, etc.).**
 - 14. publisher → publisher details.**
 - 15. shipping_method → available shipping methods with cost.**
-

3. Data Warehouse Design (OLAP)

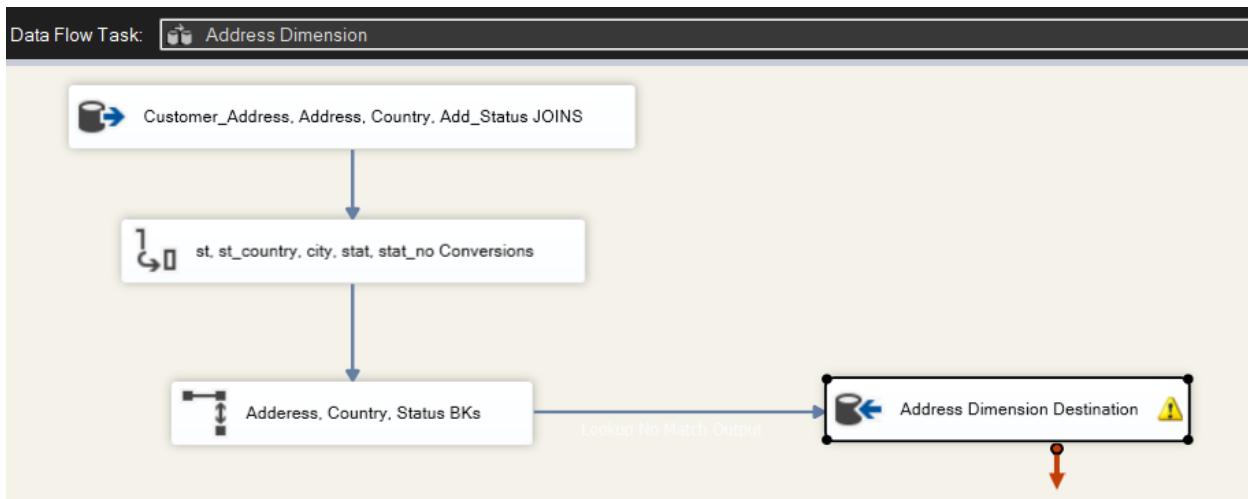
3.1 Dimensions Created

- Customer Dimension (customer_dim)**
 - Contains surrogate key, Business key, name, email, SCD attributes (start_date, end_date, iscurrent).



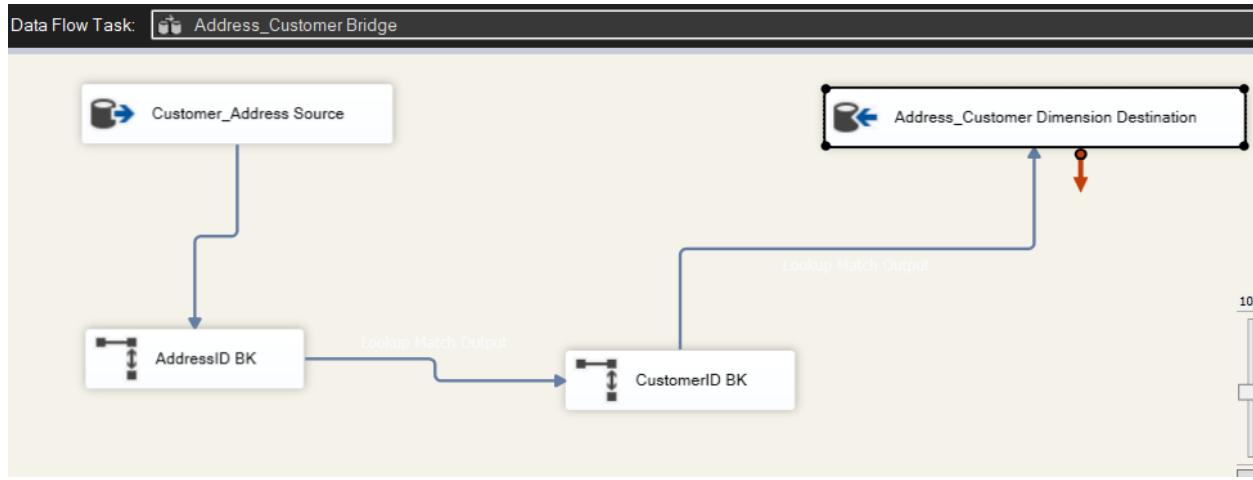
- **Address Dimension (address_dim)**

- Contains location details (street, city, country, status).



- **Customer-Address Bridge (address_cust_dim)**

- Created as a **bridge table** between customers and addresses.



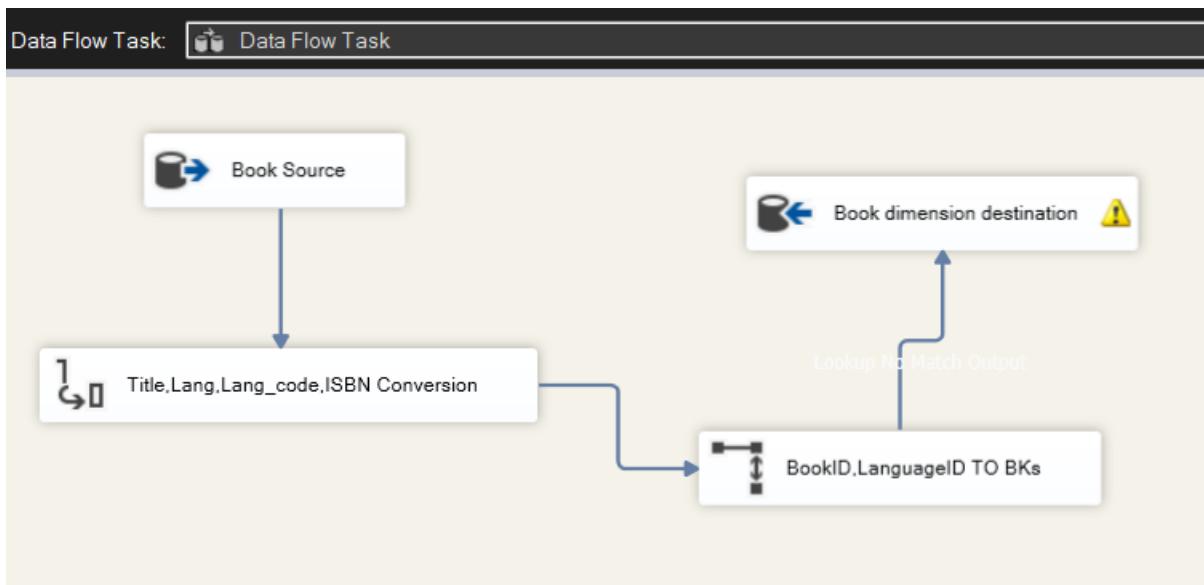
- **Author Dimension (author_dim)**

- Contains author details.



- **Book Dimension (book_dim)**

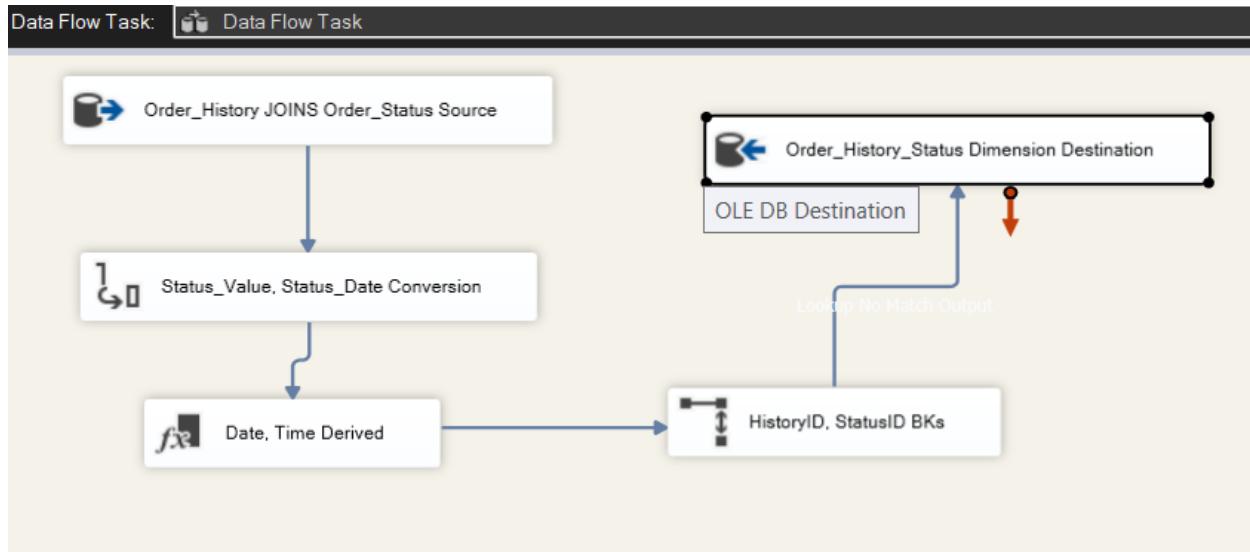
- Contains book details (title, ISBN, publisher, language, etc.).



- **Author-Book Bridge (author_book)**
 - Many-to-many bridge between authors and books.

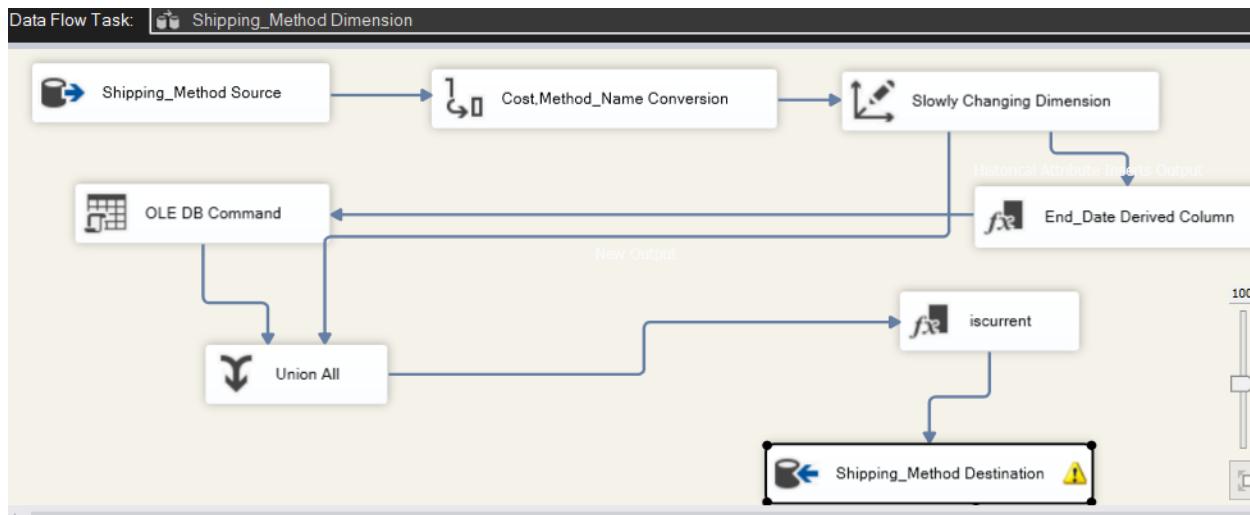


- **Date Dimension (dim_date)**
 - Split from order dates → year, month, day.
- **Time Dimension (dim_time)**
 - Extracted from timestamps → hours, minutes, seconds.
- **Order History Status (order_history_status)**
 - Implemented as a **Slowly Changing Dimension (SCD)**.



- **Shipping Method (shipping_method)**

- Also handled as **SCD**.

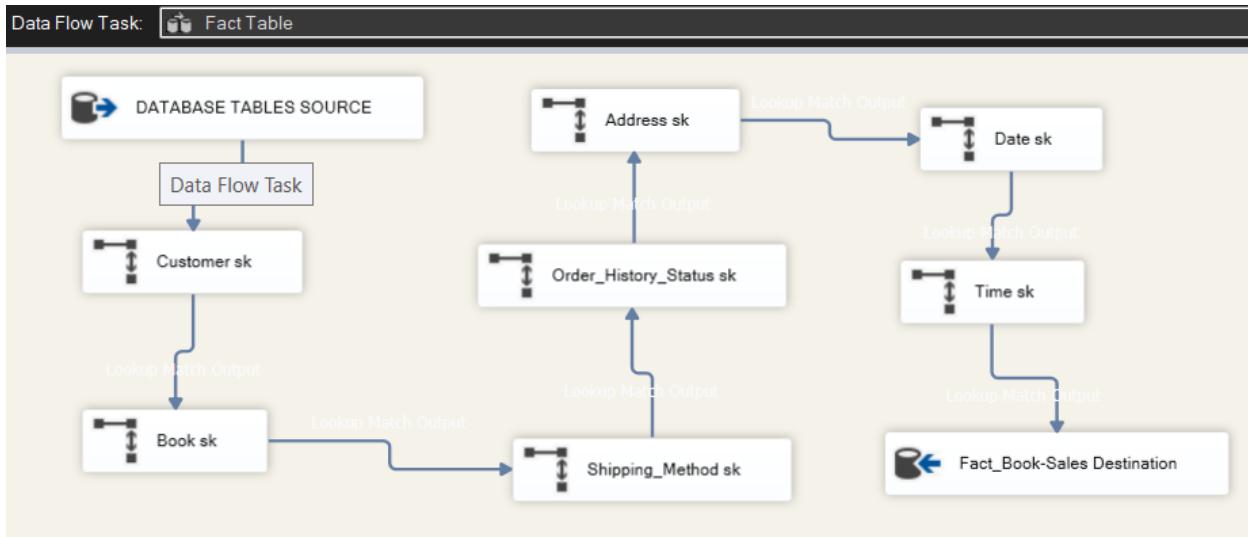


3.2 Fact Table

Fact_Book_sales

- Stores sales transactions with foreign keys:
 - Customer, Address, Book, Author, Date, Time, Order History Status, Shipping Method

- Measures:
 - Price, Cost



4. ETL Process (SSIS)

The ETL process was built using **SQL Server Integration Services (SSIS)** with the following steps:

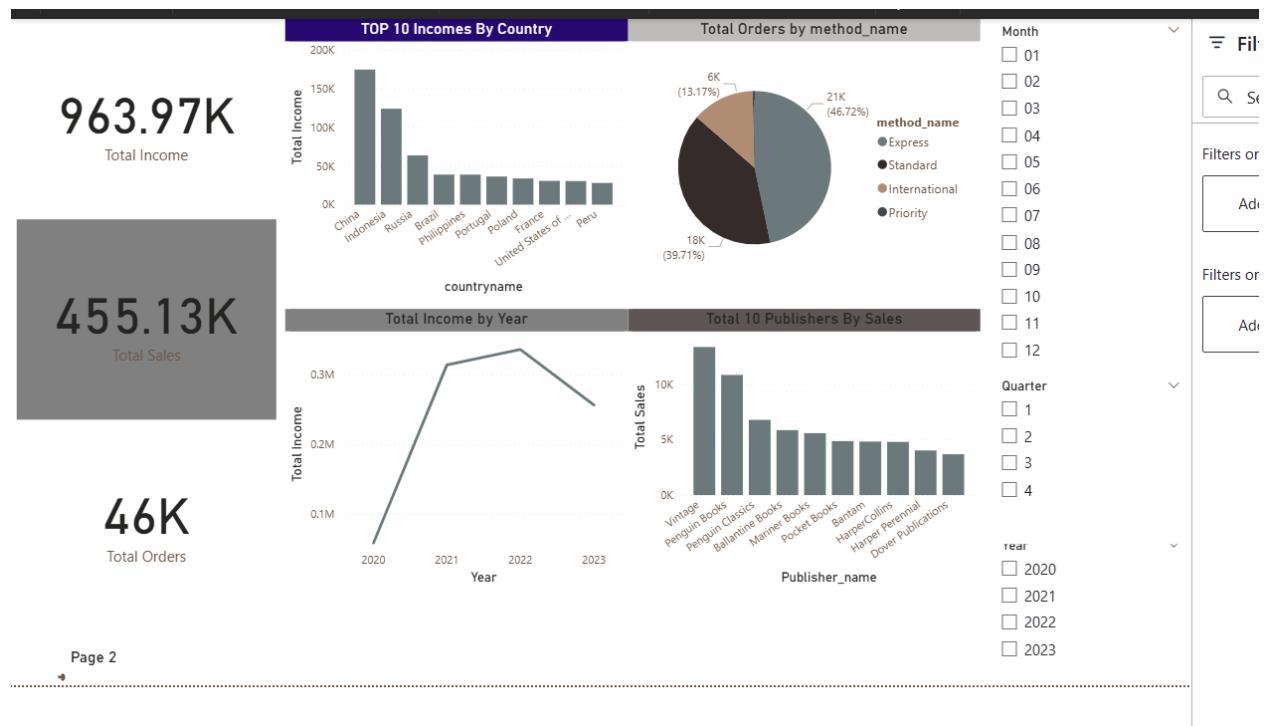
1. **Extraction:**
 - Pulled data from OLTP tables (customer, book, order_line, etc.).
2. **Transformation:**
 - **Data Conversion:** Adjusted data types (e.g., dates, string lengths).
 - **Lookup Transformations:** Connected facts to dimension surrogate keys.
 - **Split:** Extracted date & time into separate tables.
 - **Slowly Changing Dimension (SCD):** Applied to order_history_status and shipping_method.
3. **Loading:**

- Loaded into snowflake schema tables (dims, facts, bridges) in the DWH.
-

5. Power BI Dashboard

After completing the ETL process and building the data warehouse, a Power BI dashboard was developed to provide business insights.

Page 1

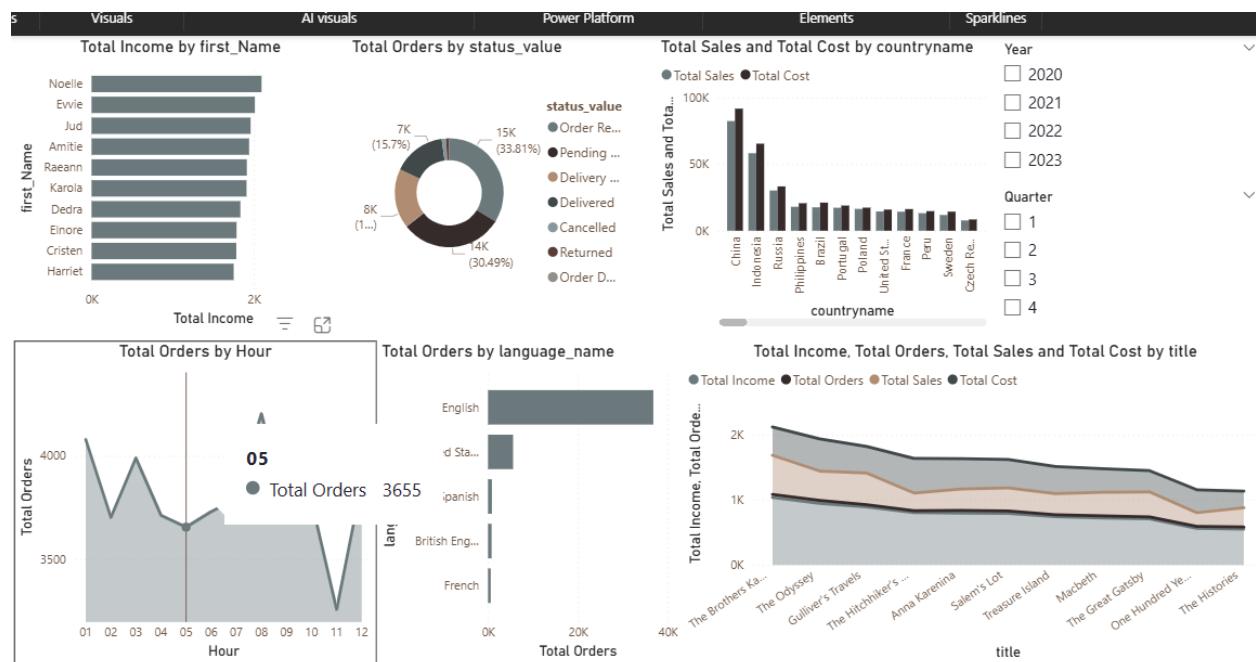


Page 1 contains a mix of KPI indicators and detailed visuals that provide an overview of income, sales, orders, shipping, and publisher performance.

- **Total Income (Card):** Shows the overall income generated, combining book sales and shipping cost.
- **Total Sales (Card):** Displays revenue generated only from book sales.
- **Total Orders (Card):** Indicates the total number of orders placed by customers.

- Top 10 Incomes by Country (Bar Chart): Highlights the countries that contribute the most income, allowing identification of the strongest markets.
- Total Orders by Shipping Method (Pie Chart): Shows distribution of customer preferences across shipping methods (Standard, Express, Priority, International).
- Total Income by Year (Line Chart): Provides a trend of income growth or decline across years, indicating performance over time.
- Top 10 Publishers by Sales (Bar Chart): Identifies the publishers whose books generate the most sales.

Page 2



Page 2 focuses on customer details, order behaviors, cost analysis, and book-level performance.

- Total Income by First Name (Bar Chart): Shows the top customers by income contribution.
 - Total Orders by Status (Donut Chart): Breaks down orders by their current status (Delivered, Pending, Cancelled, etc.), reflecting customer experience and operational flow.
 - Total Sales and Total Cost by Country (Bar Chart): Compares revenue versus costs across countries, giving insight into profitability per region.
 - Total Orders by Hour (Line Chart): Displays when orders are placed during the day, showing peak activity hours.
 - Total Orders by Language (Bar Chart): Compares demand for books in different languages.
 - Total Income, Orders, Sales, and Cost by Title (Area Chart): Analyzes book-level performance across multiple measures (income, orders, sales, shipping cost) for a detailed view of product success.
-

6. Conclusion

- The **Gravity Books DWH** successfully transformed a normalized OLTP schema into a **star schema** suitable for analytics.
- SSIS ETL pipelines automated extraction, transformation, and loading of data.
- SCD management ensured historical accuracy for dimensions.
- The **Power BI dashboard** provides clear, interactive insights for decision-making, enabling monitoring of sales trends, customer distribution, and performance of publishers/authors.