Food Trucks - Business Analysis Documentation

October 13, 2024

1 Food Trucks - Business Analysis Documentation

1.1 Introduction

1.1.1 Project Overview

This project analyzes the sales and customer data from a global network of food trucks using **Snowflake** for data storage and processing, and **Tableau** for visualization. The goal is to identify key revenue drivers, customer preferences, and regional performance to support strategic decision-making.

1.1.2 Objectives

- To analyze **POS (Point of Sale) sales data** and **customer loyalty data** for insights into sales performance across different regions.
- To integrate **geospatial data** for deeper analysis of sales relative to tourist hotspots and high-traffic locations.
- To build **interactive Tableau dashboards** for easy exploration of the insights by stakeholders.

1.1.3 Key Technologies

- **Snowflake**: Cloud-based data platform used for data storage, transformation, and querying.
- **SQL**: Used for managing data in Snowflake and creating views.
- **Tableau**: Data visualization tool used to create interactive dashboards.
- **AWS S3**: Used as a data storage layer for raw data ingestion into Snowflake.

1.2 Data Sources

1.2.1 Data Ingestion and Preparation

The project involves multiple data sources which are ingested into **Snowflake** and processed for analysis.

Data Sets and Formats

- **POS Sales Data**: Captures transaction details from food trucks.
- **Customer Loyalty Data**: Contains information about customer demographics and loyalty program participation.

- **Geospatial Data**: Includes data about tourist spots and locations relevant to food truck positioning.

Data Description

- **POS Sales Data**: Includes fields like `order_id`, `order_total`, `truck_id`, `location_id`, and `order_ts` (timestamp).
- **Customer Loyalty Data**: Includes `customer_id`, `first_name`, `last_name`, `city`, `preferred_language`, and `birthday_date`.
- **Geospatial Data**: Contains fields such as `latitude`, `longitude`, `placekey`, and `location_name`.

1.2.2 Data Dictionary

Below is a data dictionary for the key tables used in this project:

Table Name	Column Name	Data Type	Description
`order_header`	`order_id`	NUMBER	Unique identifier for each order.
`order_header`	`order_total`	NUMBER(38,4)	Total amount of the order.
`customer_loyalty`	`customer_id`	NUMBER	Unique identifier for each customer.
`menu`	`menu_item_name`	VARCHAR	Name of the menu item ordered.

1.3 Data Management

1.3.1 Snowflake Data Architecture

Database Structure

- **Database**: `frostbyte_tasty_bytes`
- **Schemas**:
  - `raw_pos`: Stores raw POS data.
  - `raw_customer`: Contains raw customer loyalty data.
  - `harmonized`: Stores processed views for analysis.
  - `analytics`: Contains views that are consumed by Tableau dashboards.

Data Organization and Storage Strategy

```
-- Create the database and schemas
CREATE OR REPLACE DATABASE frostbyte_tasty_bytes;
CREATE OR REPLACE SCHEMA frostbyte_tasty_bytes.raw_pos;
CREATE OR REPLACE SCHEMA frostbyte_tasty_bytes.raw_customer;
CREATE OR REPLACE SCHEMA frostbyte_tasty_bytes.harmonized;
CREATE OR REPLACE SCHEMA frostbyte_tasty_bytes.analytics;
```

- Database Design:
  - **Raw Data Schemas (raw_pos, raw_customer)**: Store unprocessed data directly from data sources, ensuring traceability.

- **Harmonized Schema (harmonized)**: Integrates data from multiple sources into a unified format, combining POS sales with customer demographics and geospatial data.
- **Analytics Schema (analytics)**: Contains pre-aggregated views that support fast queries in Tableau for interactive analysis.

1.3.2 Data Transformation and Modeling

Data Cleaning and Standardization Data cleaning is an essential step to ensure the accuracy of analysis. The process involves:
- **Filtering Invalid Orders**: Excludes records with negative or zero order_total.
- **Date Normalization**: Ensures consistent date formats (YYYY-MM-DD).
- **Deduplication**: Removes duplicate records to maintain unique customer entries.

```
-- Remove orders with zero or negative totals
CREATE OR REPLACE VIEW frostbyte_tasty_bytes.harmonized.valid_orders_v AS
SELECT *
FROM frostbyte_tasty_bytes.raw_pos.order_header
WHERE order_total > 0;

-- Ensure each customer record is unique
CREATE OR REPLACE VIEW frostbyte_tasty_bytes.harmonized.unique_customers_v AS
SELECT DISTINCT *
FROM frostbyte_tasty_bytes.raw_customer.customer_loyalty;
```

Data Integration: Harmonized Views Harmonized views combine data from raw_pos and raw_customer, enriched with geospatial data to provide a comprehensive dataset for analysis.

```
-- Create a view that integrates POS data with customer and geospatial data
CREATE OR REPLACE VIEW frostbyte_tasty_bytes.harmonized.orders_v AS
SELECT
    oh.order_id,
    oh.order_total,
    oh.order_ts,
    cl.customer_id,
    cl.first_name,
    cl.last_name,
    cl.city AS customer_city,
    geo.city_population,
    geo.latitude,
    geo.longitude,
    od.menu_item_id,
    m.menu_item_name
FROM frostbyte_tasty_bytes.harmonized.valid_orders_v oh
JOIN frostbyte_tasty_bytes.harmonized.unique_customers_v cl
    ON oh.customer_id = cl.customer_id
LEFT JOIN frostbyte_weathersource.onpoint_id.postal_codes geo
    ON cl.city = geo.city
JOIN frostbyte_tasty_bytes.raw_pos.order_detail od
    ON oh.order_id = od.order_id
```

```
JOIN frostbyte_tasty_bytes.raw_pos.menu m
    ON od.menu_item_id = m.menu_item_id;
```

- **Purpose**: The orders_v view aggregates key metrics such as sales amounts, customer details, and geographic information, making it ready for visualization in Tableau.

1.3.3 Data Security and Compliance

Role-Based Access Control (RBAC) RBAC ensures that data access is controlled based on user roles, protecting sensitive information while allowing users to perform their necessary tasks.

- **Defined Roles**:
  - `tasty_admin`: Manages all aspects of the database, including user roles.
  - `tasty_data_engineer`: Manages data transformation processes.
  - `tasty_bi`: Has read-only access for creating Tableau dashboards.

```
-- Create roles for access control
CREATE ROLE IF NOT EXISTS tasty_admin COMMENT = 'Admin role with full access';
CREATE ROLE IF NOT EXISTS tasty_data_engineer COMMENT = 'Data engineer role';
CREATE ROLE IF NOT EXISTS tasty_bi COMMENT = 'Business intelligence role';

-- Assign permissions to roles
GRANT CREATE DATABASE, CREATE SCHEMA, CREATE WAREHOUSE ON ACCOUNT TO ROLE tasty_admin;
GRANT USAGE, SELECT ON DATABASE frostbyte_tasty_bytes TO ROLE tasty_data_engineer;
GRANT SELECT ON SCHEMA frostbyte_tasty_bytes.analytics TO ROLE tasty_bi;
```

Data Masking and Privacy To comply with data privacy regulations, masking policies are applied to sensitive fields such as customer emails.

```
-- Create masking policy for customer email
CREATE MASKING POLICY mask_email AS (val STRING)
RETURNS STRING ->
CASE
    WHEN current_role() IN ('tasty_admin', 'tasty_data_engineer') THEN val
    ELSE '*** MASKED ***'
END;

-- Apply masking policy to the email column
ALTER TABLE frostbyte_tasty_bytes.raw_customer.customer_loyalty
MODIFY COLUMN email SET MASKING POLICY mask_email;
```

- **Purpose**: This ensures sensitive information is only visible to users with the necessary permissions, maintaining data privacy.

1.4 Analysis and Results

1.4.1 Analytical Strategy in Tableau

Focus Areas of Analysis The project uses **Tableau** to provide visual insights into several key areas: - **Regional Sales Performance**: Identify top-performing cities and regions, and understand seasonal fluctuations in sales. - **Product Preferences**: Analyze the popularity of different menu items to optimize inventory and pricing. - **Customer Behavior**: Study purchasing trends among loyalty program members, segmenting them by demographic information.

1.4.2 Detailed Dashboards

Global Sales Analysis

- **Visualization**: A **map view** displaying sales volumes across different countries, with bubble sizes representing total sales and colors indicating average order value.
- **SQL View**:

```sql
-- View to aggregate sales data by country
CREATE OR REPLACE VIEW frostbyte_tasty_bytes.analytics.global_sales_v AS
SELECT
    country,
    SUM(order_total) AS total_sales,
    AVG(order_total) AS avg_order_value,
    COUNT(order_id) AS total_orders
FROM frostbyte_tasty_bytes.harmonized.orders_v
GROUP BY country;
```

- **Purpose**: This view enables Tableau to visualize sales distributions across countries, providing insights into which regions are generating the highest revenues.

1.4.3 Technical Insights and Best Practices

- **SQL Optimization**: Use of window functions for efficient aggregations and JOIN statements to optimize query performance.
- **Dashboard Performance**: Leveraging Tableau's Hyper extracts for efficient data retrieval, enabling faster loading of dashboards.
- **Data Validation**: Implemented data validation checks to ensure data accuracy before making it available in the analytics schema.

```sql
-- Validate data integrity by checking for negative totals
SELECT COUNT(*) FROM frostbyte_tasty_bytes.harmonized.valid_orders_v
WHERE order_total < 0;
```

1.5 Conclusion

1.5.1 Summary of the Analysis

This analysis provided actionable insights into sales trends, customer preferences, and regional performance across the food truck network.

1.5.2 Key Outcomes

- `Built automated data pipelines using Snowflake, integrated with AWS S3.`
- `Developed dynamic Tableau dashboards for business stakeholders to explore.`
- `Enabled real-time analysis through direct connections between Tableau and Snowflake.`

1.5.3 Future Enhancements

- `Integrate machine learning using Snowpark to forecast sales.`
- `Use advanced geospatial analysis for optimizing truck locations.`
- `Automate data refreshes with Snowflake tasks and Tableau schedules.`

1.6 Project Link

Link to Project on Tableau Cloud: [https://prod-ca-a.online.tableau.com/#/site/sriv2930-e846d1ad58/workbooks/769125/views]

1.7 Contact Information

For further inquiries or to collaborate on similar projects, please contact:

- **Email**: [sanskarsrivastava2001@gmail.com]
- **LinkedIn**: [https://www.linkedin.com/in/sanskar-srivastava-9074541a4/]
- **GitHub**: [https://github.com/Sanskar220901]
- **Portfolio**: [https://sanskarsrivastava.com/]