Functional Specification Document (FSD)

1. Introduction

Project Overview:

In this project, we'll be analyzing the data from DummyJSON API endpoints: `/users`, `/products`, and `/carts`.

The goal is to create a Power BI dashboard to visualize product sales, user information, and shopping cart activities.

Scope:

- Extract data from the three APIs.
- Clean, transform, and analyze the data.
- Create a dashboard with visualizations (such as bar charts, line charts, pie charts, etc.) in Power BI.

2. Functional Requirements

2.1 Data Source

API Endpoints:

- **Users:** https://dummyjson.com/users
- **Products:** https://dummyjson.com/products
- **Carts:** https://dummyjson.com/carts

Data Fields:

- **Users:** id, firstName, lastName, email, gender, address, phone, etc.
- **Products:** id, name, category, price, quantity, etc.
- **Carts:** id, userId, products[], total, dateCreated, etc.

API Data Retrieval:

Each of these endpoints will be connected to Power BI to extract the data using the Web connector.

Transformations in Power Query:

- Data cleaning and transformation processes, such as removing unnecessary fields, renaming columns, and merging data across tables.

2.2 Data Flow & Relationships

Users and Carts Relationship:

- Each user has an associated shopping cart.
- The `userId` in the Carts dataset will be linked with the `id` in the Users dataset to analyze user-specific shopping behavior.

Carts and Products Relationship:

- The `products[]` field in the Carts dataset includes the IDs of the products that the user added to their cart. This can be connected to the Products dataset to analyze which products are being added the most.

Power BI Relationships:

- **Users ↔ Carts:** Using the `userId` field.
- **Carts ↔ Products:** Using the `productId` in the `products[]` field.

3. Functional Specifications

3.1 Power BI Visualizations

Dashboard Layout:

- **Main Page:** A comprehensive overview showing key metrics (total sales, number of active users, total carts, etc.).

Charts:

- **Bar Chart:** Top products by quantity sold.
- **Line Chart:** Monthly sales trends (if date information is available).
- **Pie Chart:** Product categories distribution.
- **Card/Measure Visuals:** Total revenue, total products in carts, number of unique users.

3.2 Data Transformations

Users Data:

- Filter out unnecessary columns.
- Parse and clean the address field if needed.
- Aggregate or group by user characteristics (e.g., gender, region).

Products Data:

- Clean the product names if needed (e.g., removing special characters).
- Calculate total revenue per product.

Carts Data:

- Flatten the `products[]` field to a usable format for analysis.
- Calculate cart totals, group by user or product ID.
- Add date attributes for time-based analysis (e.g., extract month from `dateCreated` if available).

4. Non-Functional Requirements

Performance:

- The Power BI dashboard should be capable of handling a high volume of data without significant performance degradation.

Data Refresh:

- The dashboard should be set to refresh the data automatically (e.g., daily or weekly).

Data Security:

- Ensure that no sensitive information (such as personal details of users) is included in the dashboard.

5. User Stories

5.1 As a Data Analyst:

- I want to connect and extract data from the DummyJSON API.
- So that I can analyze user behavior, product sales, and shopping cart activity.

5.2 As a Business Stakeholder:

- I want to see visualizations that show the best-selling products, trends over time, and the distribution of users across different regions.
- So that I can make better business decisions about inventory and marketing strategies.

6. Insights & Analysis

6.1 Key Findings from the Dashboard:

- **Age Distribution:** Most users are between 26 to 33 years old.
- **Gender Distribution:** 51% male, 49% female, almost evenly split.
- **Geographic Spread:** No anomalies; users are evenly distributed across states.
- **Total Revenue:** $3,743,551.
- **Average Order Value:** $18,907.
- **Anomalies:** High-priced products like cars make up **62% of total sales**.
- **Discount Impact:** Sales are **51-53% before discount** and **47-48% after discount**.
- **Product Availability:** 93 products are in stock, 4 are low in stock, and 3 are out of stock.
- **Ratings vs. Sales:** No clear correlation.
- **Top-Selling Products:**
  a. Family Tree Photo Frame
  b. Volleyball
  c. Pacifica Touring
  d. Rolex Datejust
  e. Baseball Ball
  f. iPhone X
  g. Selfie Lamp with iPhone
  h. Vivo S1
- **Top Categories:**
  a. Vehicles
  b. Men’s Watches
  c. Women’s Watches
- **Top States by Users:**
  - Arkansas
  - New Hampshire
  - Utah
  - Vermont
  - West Virginia
  - Delaware
  - Mississippi
  - Idaho
  - Maryland
  - Montana
  - Tennessee

7. Acceptance Criteria

The data from all three APIs should be connected and transformed correctly into Power BI.

The Power BI dashboard should display:

- Top products by quantity sold.
- Sales trends over time.
- User distribution by gender, region, etc.
- All charts should be interactive and responsive.
- Data refresh should work smoothly without errors.

8. Testing

Data Consistency Test:

- Ensure that the data extracted from the APIs matches the structure and expected values.

Visualization Test:

- Verify that the charts and visuals render correctly and reflect the correct analysis.

Performance Test:

- Test the dashboard performance with larger datasets to ensure it loads and refreshes efficiently.

9. Deployment and Maintenance

Deployment:

- The Power BI report should be published to the Power BI service for sharing with stakeholders.

Maintenance:

- Regular data refreshes should be scheduled.
- The API endpoints should be monitored for any changes or downtime that could affect data extraction.