**Project Documentation: DataSpark: Illuminating Insights for Global Electronics**

**1. Project Overview**

**Project Title**: Data Spark: Illuminating Insights for Global Electronics  
**Project Type**: Data Analytics & Insights Platform  
**Industry**: Electronics, Retail Analytics.

## Tools Used

* Jupyter Notebook
* Python, Pandas: Data cleaning and exploratory data analysis
* MySQL: Database to store and retrieve cleaned data
* Power BI: Visualization

**1.1 Objective**

The objective of **DataSpark** is to provide advanced data analytics solutions to electronics companies by analysing their retail data. The platform uses real-time insights to enhance decision-making, optimize inventory, improve customer targeting, and forecast trends within the electronics sector. The primary goal is to empower retail managers, product managers, and business analysts with actionable insights that drive sales and improve operational efficiency.

**2. Problem Statement**

Electronics retailers face several challenges in managing their supply chains, understanding customer preferences, and forecasting sales. Specific issues include:

* **Inventory management problems** due to demand fluctuations.
* **Lack of actionable insights** on customer preferences and buying behavior.
* **Inefficient sales forecasting** leading to understocking or overstocking.
* **Difficulty in analysing large volumes of retail data** that come from multiple sources.

The **DataSpark** project aims to solve these problems using a combination of data cleaning, exploratory data analysis (EDA), and powerful visualization tools.

**3. Project Scope**

**3.1 Data Collection & Integration**

Data was sourced from multiple retail databases, including sales transactions, customer reviews, product data, and inventory levels. This data was collected via APIs, ETL pipelines, and direct database access.

**3.2 Data Analytics**

The project focused on:

* **Customer Insights**: Segmenting customers based on buying patterns and demographics.
* **Sales Trends**: Analyzing seasonal trends, promotions, and product demand.
* **Inventory Optimization**: Predicting stock levels and suggesting reorder points.

**3.3 Visualization & Reporting**

Using **Power BI** interactive dashboards were created for stakeholders to make data-driven decisions.

**. Methodology and Tools**

**4.1 Data Cleaning and Preprocessing**

* **Handling Missing Data**: Imputation techniques were applied where necessary, such as mean imputation or using machine learning models for predicting missing values.
* **Removing Duplicates**: Identified and removed duplicate entries from the transaction dataset.
* **Standardization and Normalization**: Standardized categorical data (e.g., product categories, customer regions) and normalized numerical fields (e.g., sales figures) to prepare for modeling.

**4.2 Exploratory Data Analysis (EDA)**

* **Descriptive Statistics**: Summarized data with mean, median, and standard deviation to understand the general distribution.
* **Visualization**: Visualized data using histograms, box plots, and scatter plots to identify outliers, trends, and correlations between variables.

**4.3 Data Management Using SQL**

* **Database Design**: Designed SQL databases to store all the pre-processed and cleaned data.
* **SQL Queries**: Efficient queries were written to extract insights and perform aggregations (e.g., total sales by region, product sales over time).

**4.5 Data Visualization**

* **Power BI**: Used Power BI to create dashboards for executive-level insights on sales trends, customer behaviour, and product performance.

**5. Key Insights of the Project:**

* Customer Segmentation: Identified key customer groups based on purchasing behaviour.
* Sales Performance: Analysed trends over time and by product/store.
* Product Analysis: Determined top-selling products and categories.
* Store Analysis: Evaluated store performance and identified high-growth regions.

## 6. Results and Insights from the Project:

1. Customer Analysis indicates that male customers significantly outnumber the female customers with a slight difference.
2. United States customer dominance is higher compared to United Kingdom, Canada, Germany and Australia. Least Customers were from Netherlands, France and Italy.
3. Charles Collins ,Ollie Davis and Castimir Pejic were top 3 customers contributing to sum of revenue.
4. Customer purchase frequency, how many times each customer have purchased has been analysed.
5. Sales Analysis by Category indicates that Computers were the top category by the Quantity sold, Tv and video were the least sold.
6. Sales Analysis by Currency shows that Sum of profit for USD currency was 68.01% , followed by: EUR: 16.96% GBP: 14.35% CAD: 0.68%
7. Sales analysis by Trends in order Dates demonstrates that Combined revenue and Sales from 2019 and 2020 is higher than the total revenue from the other Period of year.
8. Store analysis by country and sales proves that US has high no of stores by 66.93% compared to Australia with 17.07% . Stores with Store Key “5” contribute to the highest sales and revenue.
9. Contoso product has the Highest unit price and Adventure work products has the least unit price.