Shriram Mantri Vidhya Nidhi Info-Tech Academy

E-commerce Data Analysis

GUIDED BY :

PRADEEP TRIPATHI

MENTORED BY :

ANIRUDDHA PARTE

Presented By :

Saurabh Khandebharad
Gaurav Mankar
Abhilash Dable
Darshana Chavare
Ketan Sawant
Umair Hashmi
Pratiksha Yerne
Akash Sadvilakar
Rajdev Yadav
Kush Yadav

Objective:

- The project aims to analyze the E-commerce store data and get key insights through visualizations so that the Management Team can make better decisions.

KAGGLE DATASET:

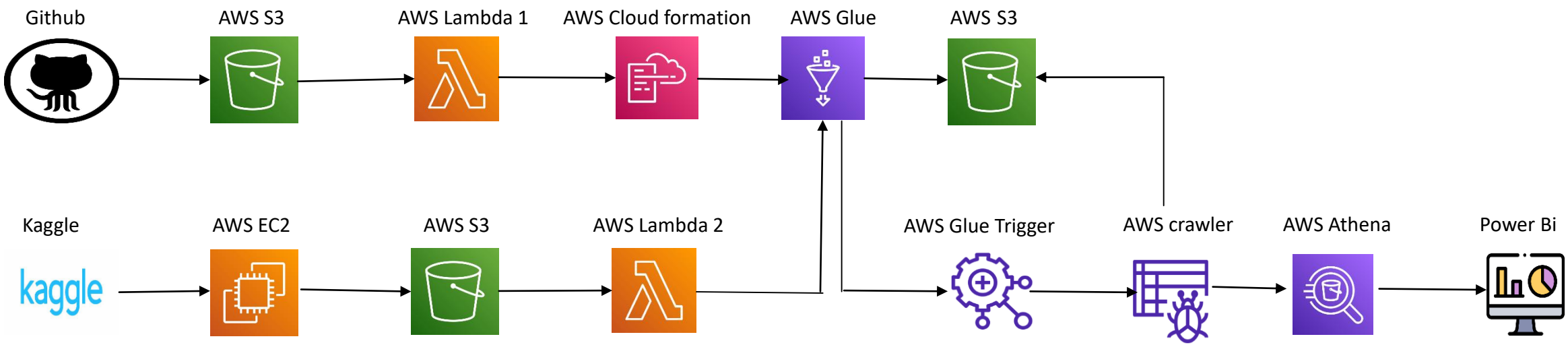
https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store

Tools & Technologies

Cloud Frameworks :-

- AWS S3
- AWS Ec2
- Apache PySpark
- AWS Athena
- AWS Crawler
- AWS Glue
- AWS Lambda
- AWS Cloud Formation
- Power Bi

Architecture:



Dataset Info

- This file contains behavior data for November 2019 month from a large multi-category online store.
- Each row in the file represents an event. All events are related to products and users.
- Size:- 8GB

- Schema before Transformation:- (no of features 8)

  Event Time, Event Type, Product_Id, Category_Id, Category_Code, brand price, User_Id, User_Session

- Schema before Transformation:- (no of features 14)

  “Event_Type", “Product_Id", “Category_Id", "brand", "price", “User_Id", “User_Session", “Category", “Subcategory", “Event_Date", “Event_Time(UTC) “, “Cart", “View", “Purchase”

Workflow:

1. Github: The transformation script is pushed to Github using Github Actions.
2. S3: The .py file containing the transformation script is stored in an S3 bucket.
3. Lambda1: A Lambda function is triggered to create a CloudFormation template (CFT).
4. AWS CloudFormation: The CFT is created to define the infrastructure for the Glue job.
5. AWS Glue Job: The Glue job is created to perform the data transformation. The .py file containing the transformation script is used by the Glue job to transform the data.
6. Kaggle: The data is extracted from Kaggle and uploaded to S3 using an EC2 instance.

Workflow continued..

7. Lambda2: A Lambda function is triggered when the data is uploaded to S3. The Lambda function runs the Glue job created by the CFT.

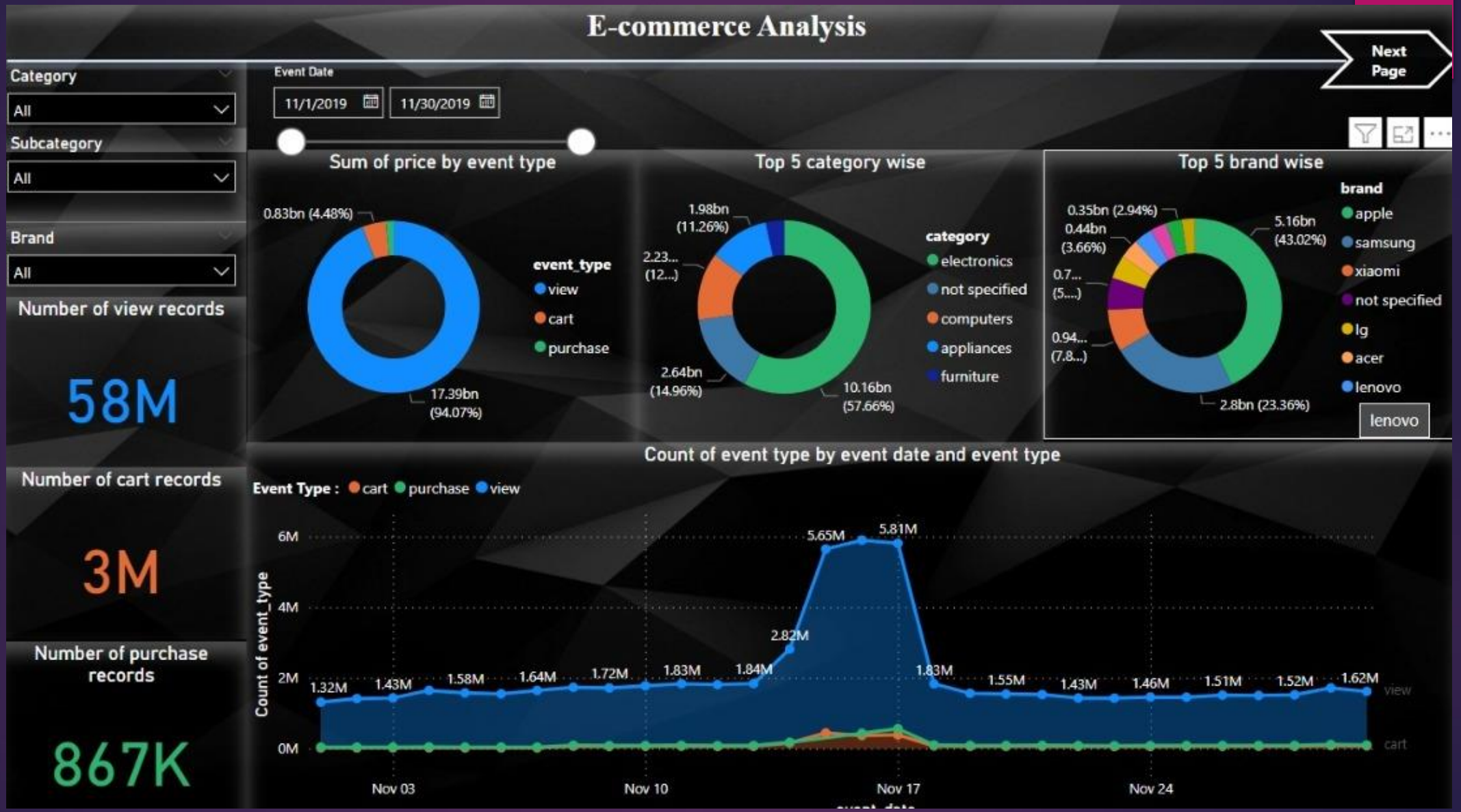
8. AWS S3: The transformed data is stored in an S3 bucket.

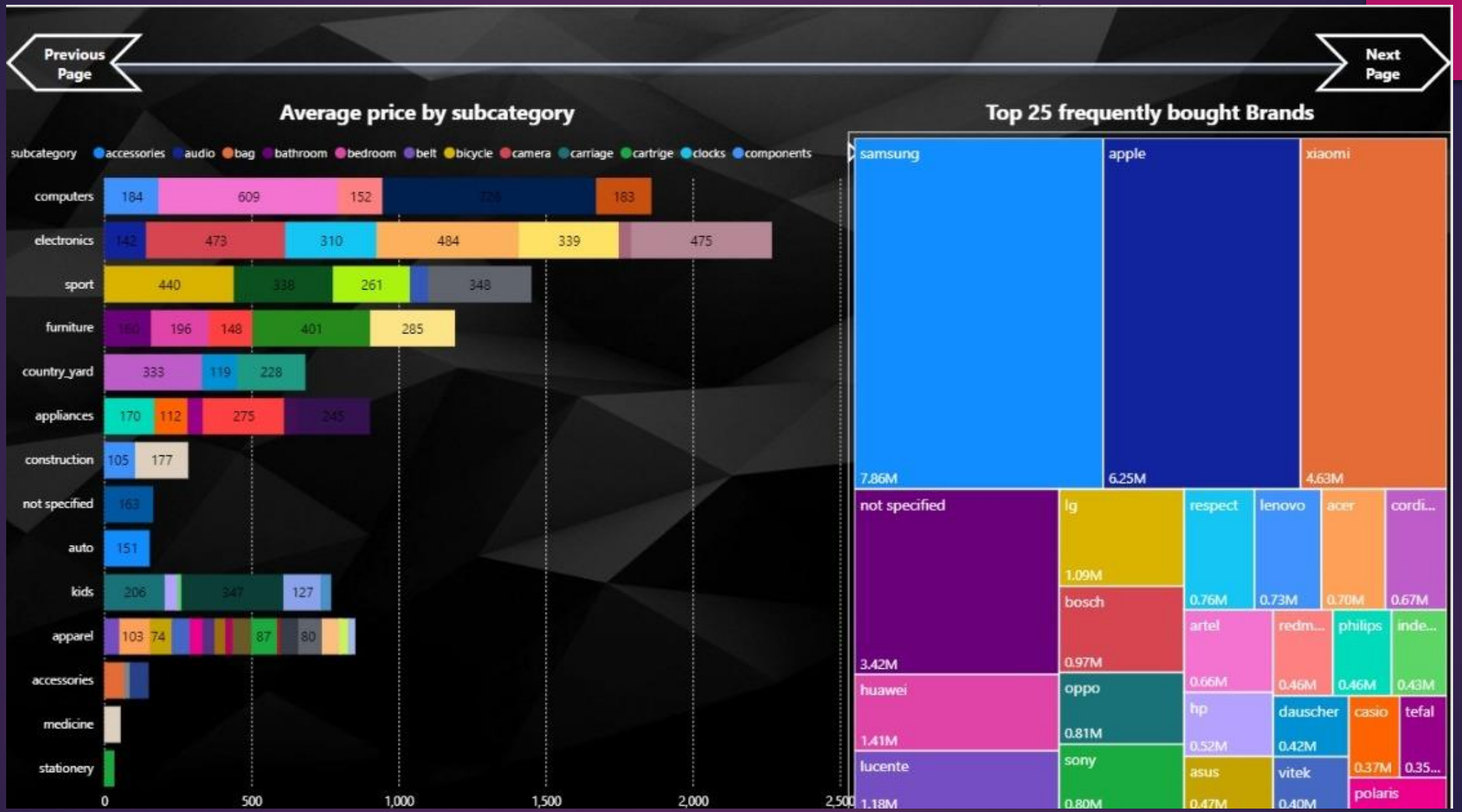
9. Glue Trigger: A Glue Trigger is set up to automatically run the Crawler and update the table in the Glue Data Catalog with the newly transformed data.

10. AWS Crawler: A Crawler is triggered to automatically crawl the transformed data and create a table in the AWS Glue Data Catalog.

11. AWS Athena: The data can be analyzed using Athena, which is an interactive query service that makes it easy to analyze data in S3 using standard SQL.

12. Power BI: The data can be visualized using Power BI, which is a business analytics service provided by Microsoft.


E-commerce Analysis
Next Page
Category
All
Subcategory
All
Brand
All
Event Date
11/1/2019
11/30/2019
Number of view records
58M
Number of cart records
3M
Number of purchase records
867K
Sum of price by event type
0.83bn (4.48%)
17.39bn (94.07%)
event_type
view
cart
purchase
Top 5 category wise
1.98bn (11.26%)
2.23... (12...)
2.64bn (14.96%)
10.16bn (57.66%)
category
electronics
not specified
computers
appliances
furniture
Top 5 brand wise
0.35bn (2.94%)
0.44bn (3.66%)
0.7... (5....)
0.94... (7.8...)
5.16bn (43.02%)
2.8bn (23.36%)
brand
apple
samsung
xiaomi
not specified
lg
acer
lenovo
lenovo
Count of event type by event date and event type
Event Type : cart purchase view
Count of event_type
6M
4M
2M
0M
1.32M
1.43M
1.58M
1.64M
1.72M
1.83M
1.84M
2.82M
5.65M
5.81M
1.83M
1.55M
1.43M
1.46M
1.51M
1.52M
1.62M
view
cart
Nov 03
Nov 10
Nov 17
Nov 24


Previous Page
Next Page
Average price by subcategory
subcategory
accessories
audio
bag
bathroom
bedroom
belt
bicycle
camera
carriage
cartrige
clocks
components
computers
electronics
sport
furniture
country_yard
appliances
construction
not specified
auto
kids
apparel
accessories
medicine
stationery
0
500
1,000
1,500
2,000
2,500
Top 25 frequently bought Brands
samsung
7.86M
apple
6.25M
xiaomi
4.63M
not specified
3.42M
lg
1.09M
respect
0.76M
lenovo
0.73M
acer
0.70M
cordi...
0.67M
bosch
0.97M
artel
0.66M
redm...
0.46M
philips
0.46M
inde...
0.43M
huawei
1.41M
oppo
0.81M
hp
0.52M
dauscher
0.42M
casio
0.37M
tefal
0.35...
lucente
1.18M
sony
0.80M
asus
0.47M
vitek
0.40M
polaris

Thank You!