

Amazon Data Analysis Report

Introduction:

This report presents an innovative predictive model designed to categorize Amazon transactions into 3 state;

- Completed
- Cancellation
- Pending

During the Data exploration and cleaning, I discovered multiple insights:

1. Most sales came from the electronics category, giving approximately 48% of revenue.
2. The top three products purchased from the stores are smart watch, smart phone and laptop
3. Miami and Chicago being the country with the highest purchase of electronics
4. Books isn't performing well and has a very slow and limited number of sales
5. The key issue in this dataset is the Transaction status which has;
 - 35.2% Completed
 - 30.8% Cancelled
 - 34.0% Pending

The major cause of this decline in completed transaction comes from the payment method, majorly from debit card, credit card and the gift card section. Top three cities that uses this above listed payment methods are Houston, New York and Miami

Methodology:

This section covers data collection from Amazon transactions, preprocessing steps, and the machine learning model development.

This particular dataset was gotten from Kaggle, data cleaning and exploration was done with python, no null values and duplicates was found, it contained 249 rows and 11 columns. Filtering was done to streamline analysis

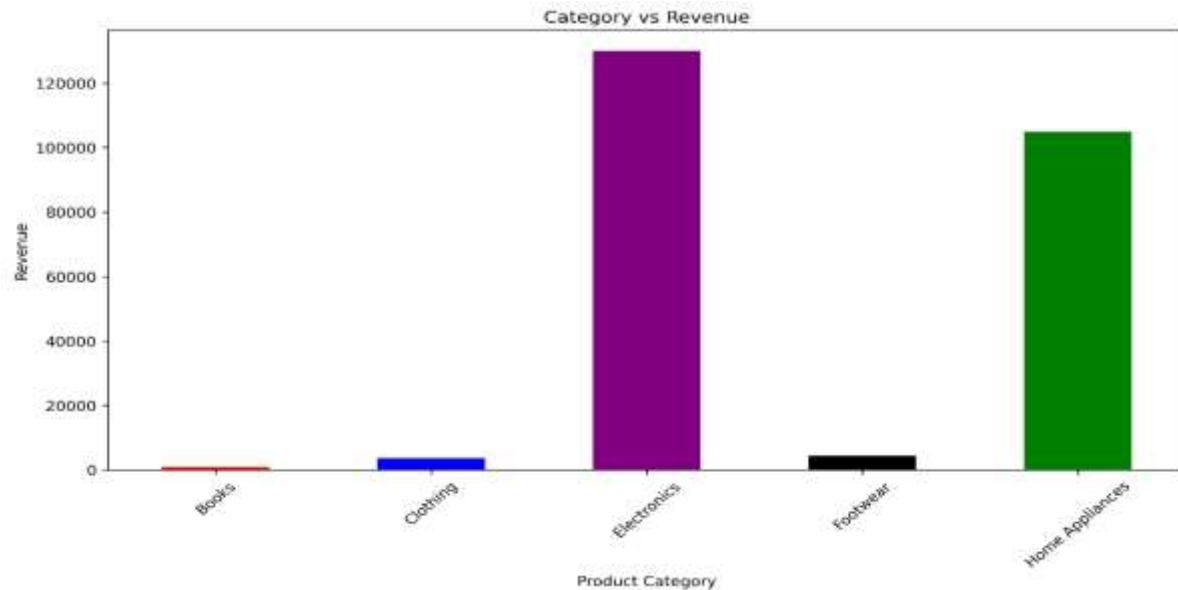
For predictive model building, Random Forest Classifier which is one of the best for predicting data that have to do with classification. Hyper-parameter tuning was tested but it furthered reduced the model's accuracy score to 91%, I had to go ahead with the models score with score of 96%

Amazon Data Analysis Report

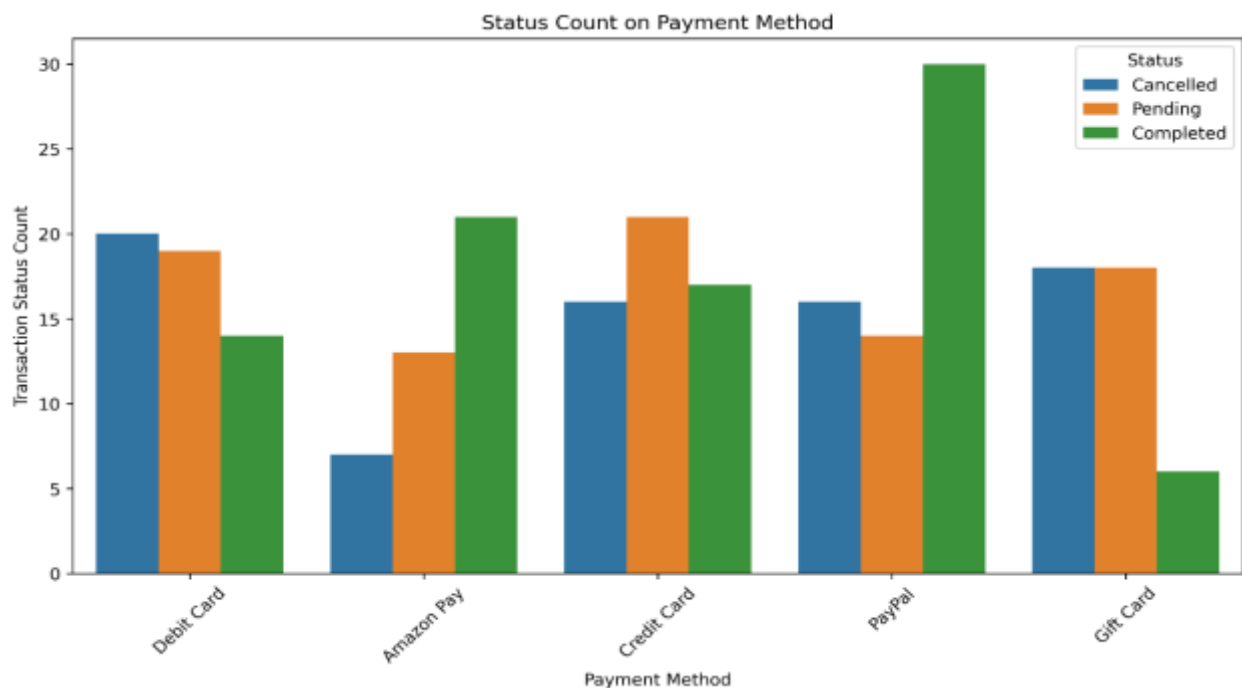
Exploratory Data Analysis:

Different visual analysis was explored on the filtered part of this dataset and this was done using matplotlib and seaborn both which are python libraries:

1. PRODUCT CATEGORY VS TOTAL SALES:

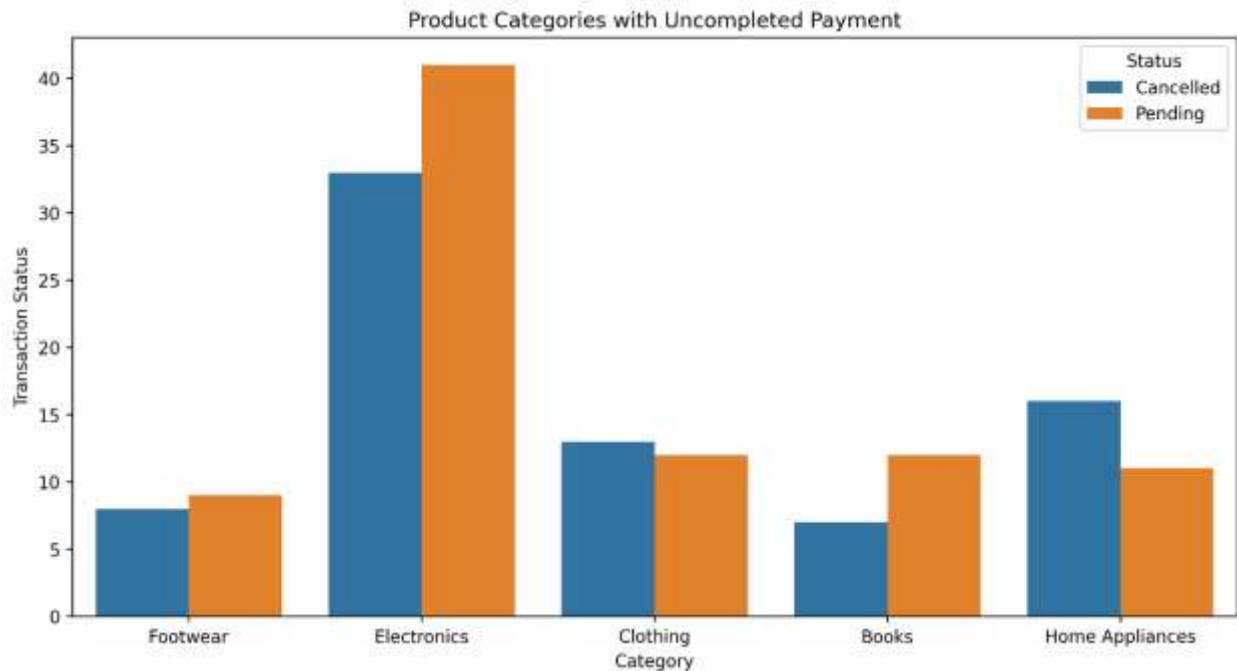


2. TRANSACTION STATUS VS PAYMENT METHOD:



Amazon Data Analysis Report

3. PRODUCT CATEGORIES WITH UNCOMPLETED PAYMENTS:



An exploration app is made on this particular dataset, which can be used to pick and explore options in the categorical data.

Deployment & User Interaction:

Deployment of both the visualization and model itself was done using streamlit.

The Model App: I used input variable which consist of;

- Product
- Category
- Total Sales
- Customer Location
- Payment Method

And all this is to ensure the model is able to predict the possible outcome of the transaction status

The Plot & Visuals App: This app is built based on the categorical columns to assist in the distribution and filtering of data. It has an inbuilt filter function which can help users segregate and clearly monitor between all the variables

Amazon Data Analysis Report

Conclusion & Impact:

The major issue is the incomplete transaction and majorly from the debit card, credit card and gift cards. This model can indicate potential customers likely to face this challenge, so alternatives like paypal and amazon pay which has high completed rates can be suggested. Therefore boosting revenue by at least 25% for the company.

REFERENCES:

- Dataset (Kaggle.com)
- Data cleaning and model building (python)
- App development (visual studio code)

Thanks for the read

Jubril ifekoya Babatunde.