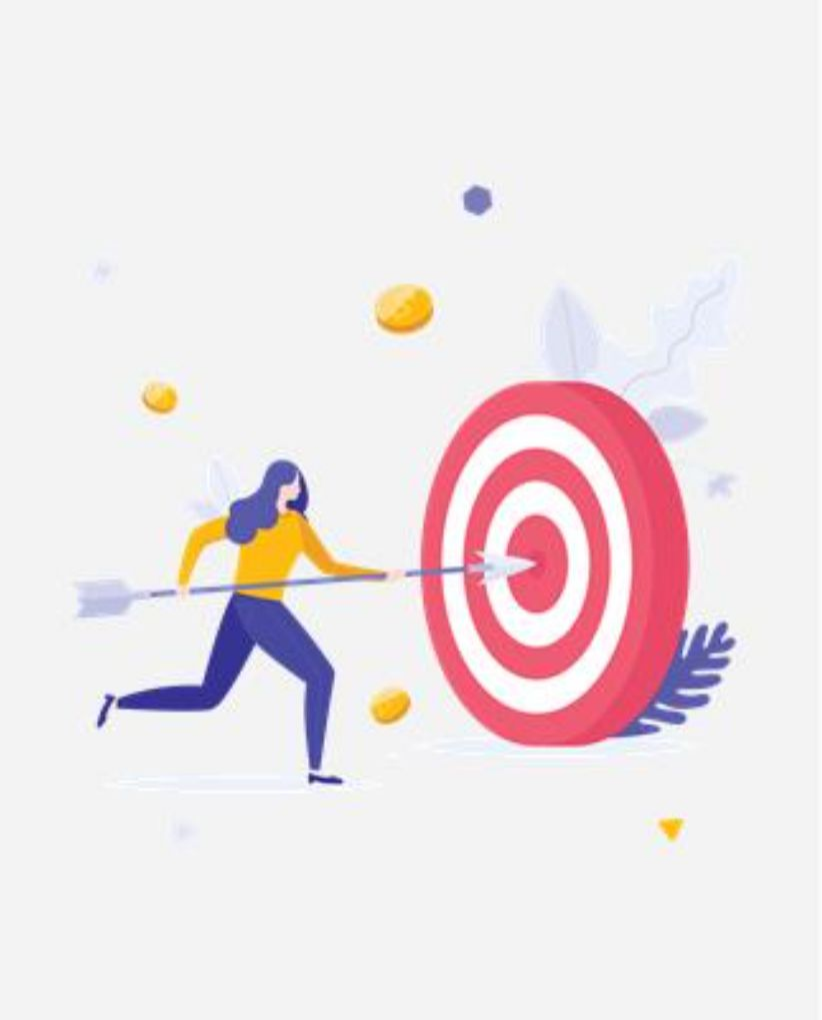
Project Problem Statement

Amazon Product Review Analysis



Project Objective

This project covers your skills in using various aspects of data analysis tools effectively.

Your solution should include a good analysis of the data and make use of the best approach for forecasting.

A good solution has a sound application of programming and algorithmic knowledge that matches the given problem statement.

Use your modular coding and analytical skills for this project.

Amazon Product Review Analysis

The year was 1994 when Jeff Bezos launched Amazon from his garage. In 1995, the first product launched by Amazon was a book in 50 states and in 45 countries within 30 days. (Oberlo 2021)

Within 26 years, Amazon became the world's largest online retailer and a household name. The Amazon name has become synonymous with online shopping and continues to grow by developing new products, acquisitions, and numerous service offerings to enlarge the customer base.

Nowadays, almost 150.6 million people turn to the Amazon app for most everything. Several types of research have proven that customers trust Amazon. (Statista 2019) On average, the small and medium-sized businesses located in the USA sell more than 4,000 items per minute (Amazon 2019), which leads to millions of product reviews on Amazon.

Amazon Product Review Analysis (contd.)

Reviews tell which products and features are trending, what's in demand, what's no longer relevant, how products and competitors are doing, and much more.

It's observed that a significant number of shoppers look at product reviews before they make a purchase. Survey results show that positive product reviews are a key factor for purchasing by 57 percent of Amazon buyers. (Statista, 2019)

As product reviews are often the deciding factor for many customers, it’s very important to have a well-automated system for monitoring them.

Amazon Product Review Analysis (contd.)

The traditional manual process of Amazon product reviews is time-consuming and inefficient when millions of reviews are being posted all the time. It doesn’t show trends or patterns over time, and it's tough to understand customer sentiment towards any product or its delivery.

Review analysis must adjust dynamically to the changing trend.

Case Study: Amazon Product Review Analysis (contd.)

Thomas, a global market analyst, wishes to develop an automated system to analyze and monitor an enormous number of reviews. By monitoring the entire review history, he will analyze tone, language, keywords, and trends over time to provide valuable insights that
increase the success rate of existing and new products and marketing campaigns.

PROJECT

Consider both of the scenarios mentioned in the following slides:

Scenario 1: Inventory Optimization and Demand Forecasting

Optimize inventory management by identifying the product categories (clustering as an outcome of text processing) on the customer review data. Predict what kind of products could be in demand (Time Series Analysis).

PROJECT

Scenario 2: Customer Retention and Sentiment Forecasting

Customer retention strategy through feedback analysis (customer classification and clustering as an outcome of analyzing the review text). Trend and seasonality analysis to predict how frequently a particular category of customer would shop in the future. (Time Series Analysis)

Time Series component: Trend, Seasonality Analysis to predict how frequently this the customer would buy new products.

PROJECT

Case Study: Amazon Product Review Analysis (contd.)

Consider **any one** of the previously stated scenarios and help Thomas build the required automated product review analyzing system.

Show your ability to tackle problems and find your own solutions which include the ability to build a model. Use the Data Science tool stack.

PROJECT

Amazon Dataset

1. Click the Base dataset to visit the website.

2. Click **5-core** and **ratings only** present on the webpage of the **Amazon product data** to download the relevant review datasets related to the chosen category as shown in the image given in the next slide.

3. Scroll down the webpage and click on **metadata** to download the data-related descriptions, price, sales-rank, brand info, and co-purchasing links of different categories of Amazon products.

The dataset contains product reviews and metadata from Amazon, including 142.8 million reviews spanning from May 1996 through July 2014.

Amazon Dataset (contd.)

Books	5-core (8,898,041 reviews)	ratings only (22,507,155 ratings)
Electronics	5-core (1,689,188 reviews)	ratings only (7,824,482 ratings)
Movies and TV	5-core (1,697,533 reviews)	ratings only (4,607,047 ratings)
CDs and Vinyl	5-core (1,097,592 reviews)	ratings only (3,749,004 ratings)
Clothing, Shoes and Jewelry	5-core (278,677 reviews)	ratings only (5,748,920 ratings)
Home and Kitchen	5-core (551,682 reviews)	ratings only (4,253,926 ratings)
Kindle Store	5-core (982,619 reviews)	ratings only (3,205,467 ratings)
Sports and Outdoors	5-core (296,337 reviews)	ratings only (3,268,695 ratings)
Cell Phones and Accessories	5-core (194,439 reviews)	ratings only (3,447,249 ratings)
Health and Personal Care	5-core (346,355 reviews)	ratings only (2,982,326 ratings)
Toys and Games	5-core (167,597 reviews)	ratings only (2,252,771 ratings)
Video Games	5-core (231,780 reviews)	ratings only (1,324,753 ratings)
Tools and Home Improvement	5-core (134,476 reviews)	ratings only (1,926,047 ratings)
Beauty	5-core (198,502 reviews)	ratings only (2,023,070 ratings)
Apps for Android	5-core (752,937 reviews)	ratings only (2,638,172 ratings)
Office Products	5-core (53,258 reviews)	ratings only (1,243,186 ratings)
Pet Supplies	5-core (157,836 reviews)	ratings only (1,235,316 ratings)
Automotive	5-core (20,473 reviews)	ratings only (1,373,768 ratings)
Grocery and Gourmet Food	5-core (151,254 reviews)	ratings only (1,297,156 ratings)
Patio, Lawn and Garden	5-core (13,272 reviews)	ratings only (993,490 ratings)
Baby	5-core (160,792 reviews)	ratings only (915,446 ratings)
Digital Music	5-core (64,706 reviews)	ratings only (836,006 ratings)
Musical Instruments	5-core (10,261 reviews)	ratings only (500,176 ratings)
Amazon Instant Video	5-core (37,126 reviews)	ratings only (583,933 ratings)

Preparing Data for Analysis

Here are the data pre-process requirements for the Amazon dataset:

- Reusability: Collecting a real dataset is very costly and time-consuming. While analyzing data, there can be a chance of losing the state of data. You should take your data backup before you start the analysis for future use.
- Free from the anomaly: Re-sample or perform required processing to remove disintegration from data.
- Missing value treatment: Make your data free from missing values.
- Validity: Check against the checklist of all assumptions required to meet in data before you start the analysis.
- Reliability: Perform required data formatting and editing to make data fit for analysis.

Sentiment Analysis Using NLP

Here is the required data analysis for the Amazon dataset:

- Natural Language Understanding: Convert a large set of text into more formal representations, such as first-order logic structures that are easier for the computer programs to manipulate notations.
- Information Extraction: Extract structured information from unstructured information.
- Sentiment Analysis: Analyze the attitude or emotional state of the customers from their posted review texts. Furthermore, make sure your work should follow a minimum of three types (Positive, Negative, Neutral) of customer sentiments.

Sentiment Analysis Using NLP (contd.)

Perform product and customer segmentation using sentiment analyzed data into:

- loyal customers vs. customers who will churn
- satisfied vs. dissatisfied customers
- products that may be good recommendations to a customer vs. not
- psychographics
- attitudes towards specific products

Sentiment Time Series Forecasting

Combining time series analysis with natural language processing, we're able to show how the sentiment of unstructured text data changes over time, as well as use it to predict future data trends.

Thank you!