

Sentiment Analysis on NFCU App Reviews

NAVYA MADDU

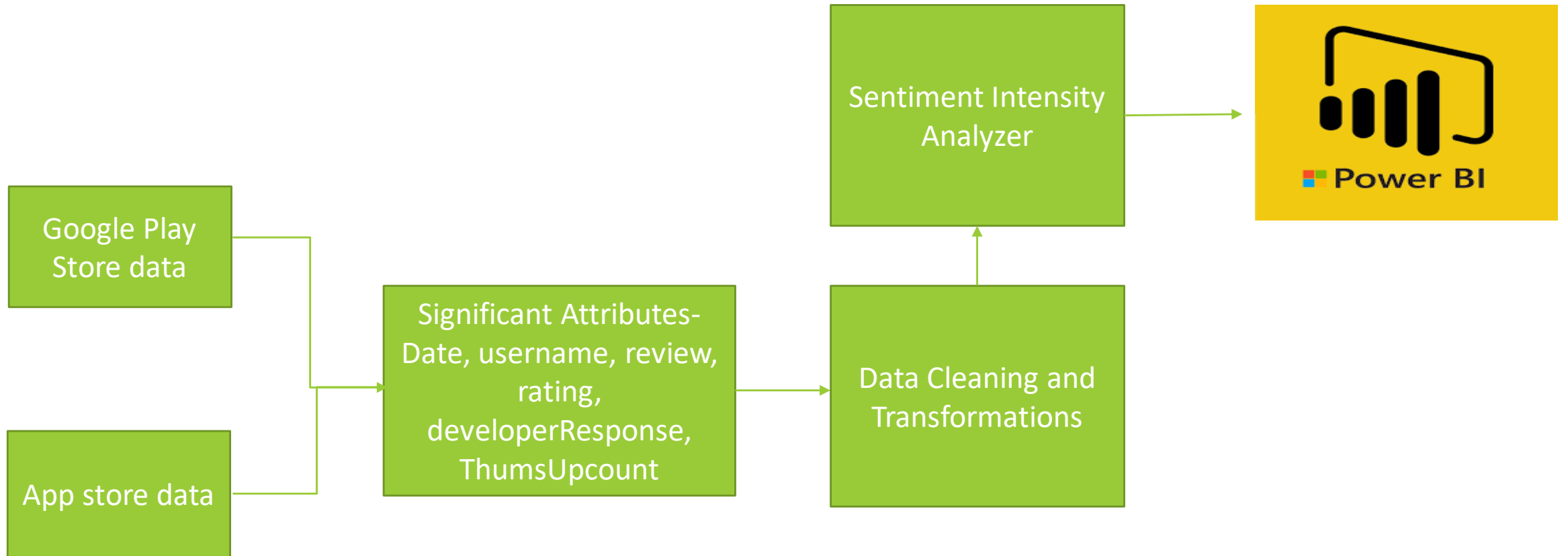
Problem Statement

- Analyze Navy Federal App reviews both on App store and Google play store to improve app experience, customer satisfaction and helping product teams to make better decisions.

Project Objectives

- Data Extraction
- Data Cleansing & Transformations
- Exploratory Data Analysis
- Performing Sentiment Analysis
- Develop an Interactive Dashboard

SYSTEM DESIGN



Data Extraction

Extracted Navy Federal App reviews from Google play Store and App store using Google Scraper and Appstore Scraper.

	Google Play Store	App Store
Total reviews	26056	4000
Total columns	10	8
Column Names	at, content, repliedAt, replyContent, reviewCreatedVersion, reviewId, score, thumpsUpCount, UserImage, UserName	Date, isedited, rating, review, title, username, thumpsUpCount, developerresponse
Date range	2011 July to 2021 July	2010 May to 2021 July
Total Responses	5420	1294

Tools Used:

- Python
- Pandas
- NLTK
- Jupyter Notebooks
- PowerBI Desktop

Data Cleansing & Transformations

- ❑ Special characters and new line characters were removed
- ❑ Performed data validation steps like duplicate checks, valid dates, data type conversions
- ❑ Required and common attributes from both data sources are extracted and combined into a single data frame.
- ❑ Reviews that are less than 3 words were removed

Sentiment analysis

- ❑ Sentiment analysis is mining of text which extracts subjective information helping business to understand social sentiment of their brand.
- ❑ SentimentIntensityAnalyser from nltk package is used to classify reviews into categories.
- ❑ The algorithm takes review text as input and returns positive, negative, neutral and compound scores.
- ❑ Compound scores are then used to categorize the review

Interactive Dashboard

New columns/measures created:

Brand : If length of reviewId is less than 5, brand is categorized as iOS, else Android.

Response Rate : Total number of responses / Total number of reviews (22.5%)

Sentiment Review:

If compound score is greater than or equal to 0.5 → Positive

If compound score is between 0.25 and 0.5 → Satisfactory

If compound score is between 0 and 0.25 → Unsatisfactory

If compound score is between -0.5 and 0 → Negative

If compound score is between -1 and -0.5 → Extremely Negative

Observations

Developers responded to 22.34% of reviews

Major spikes in Reviews during April 2019, October 2019, July 2020, November 2020 and May 2021.

Reviews from “A google user” received more likes.

Out of 30,056 reviews, 13715 reviews are Positive while 2030 are extremely negative.

Future scope

Advanced data cleansing like Stemming and Lemmatization.

Stemming – Process of reducing words to their base or root form.

Stop word removal- Most common words in a language like 'is', 'as', 'the', 'and'.

These techniques reduce lot of useless features, thereby giving us better results.

Thank You