Capstone Project 1: Project Proposal

Problem to Solve

Recognized by its famous swoosh logo, Nike has firmly established its place as a leader in sportswear innovation and as a marketing powerhouse. The company understands that in order to remain at the top, it has to build deep relationships with its members and communities online. Launched in 2015, the SNKRS app is Nike's main channel for sneakerheads to lay claim to its high-demand, limited release shoes. The present incarnation of the app allows Nike's most-engaged customers to acquire exclusive products and share their win on the app. Nike is using member-created content to showcase real people using its products – an effective marketing tactic leveraged by the clout of social-media influencers who can increase engagement among the brand's millions of more casual shoppers.

In a crowded space where well-established brands compete for buyers' attention, authenticity has proved to be one of the hallmarks customers look for when shopping digitally. Contrary to popular belief, apps with five-star ratings can actually be less influential in persuading people to download and engage with them, likely due to today's skeptical consumers' too-good-to-be-true sensibilities. Having a relatively few less-than-perfect reviews may decrease the apps average star rating, but often shows a more complete picture, making the service the app provides more real, trustworthy, and genuine to the brand's valued members.

Casually scrolling the reviews posted on Nike's SNKRS app, it's clear that some members love the app while others are quick to point out its pain points, mostly stemming from bots taking over high-profile sneaker launches. A sneaker bot is an automated script used to speed up the checkout process when purchasing products, mostly limited items, online. This trick is used by people who weaponize an army of bots to try and cheat a sneaker release. When there is strong hype surrounding the launch of a high-collaboration shoe, inventory can instantly sell out and catapult resale markups into the thousands-of-dollars range. This outcome produces a high profit margin for resellers and, at the same time, disappoints sneakerheads who receive an "L" when they lose out on a high-demand launch release.

Because Nike wants to be able to better connect to its members in a more direct way, the brand needs to understand the types of experiences users are having on the SNKRS app. Having an overall high rating on the app doesn't provide the full story of its members' experiences; instead, it's worthwhile to transform the goldmine of reviews accured on the app into valuable insights that can drive actionable business decisions.

Thus, the questions to answer and problems to solve for this first capstone project include:

- **Does the quantity of the reviews change surrounding the release of a high-profile launch?**
- **What topics/themes/trends occur most frequently across all reviews?**
- **Which words ("fair" or "bot") embedded in the reviews predict a positive or negative rating?**
- **Which topics/themes/trends predict a positive review and which predict a negative review?**
- **Does the mention of "deleting the app" or "not entering again" predict a negative review?**

The problem to solve for the second capstone project:

- **How does social media sentiment predict resale price of high-profile shoes on StockX?**

Client

NIKE, Inc. is a global sportswear company that *designs and delivers innovative products, experiences, and services to inspire athletes*. Founded in 1964 by an American track athlete and his coach, Nike has grown to become the most recognized sportswear brand in the world. The company first began by selling track shoes constructed from a waffle iron sold from the back of an automobile at local track meets. Today, the sportswear giant releases exclusive sneakers paired with high-profile collaborations, selling out in a matter of minutes all from a digital app.

At the start of 2020 and in the wake of the COVID-19 global pandemic, the sportwear brand Nike unleased its new strategy called "Consumer Direct Acceleration," which is a strategic plan to sell more products directly to customers through its website and digital platforms. Digital sales are crucial to Nike's future revenue goals. In fact, roughly 30% of Nike's revenue in its last quarter (Q4 2020 earnings) came from digital sales, an increase of 75% year-over-year. The sportswear brand recognizes the giant opportunity that the digital landscape provides, and it's investing heavily in technology and data analytics to offer a more personalized experience to its loyal consumer base.

Focusing in on the SNKRS app, Nike will find value in being able to pinpoint which topics its members are talking about that are embedded within the app's 479K reviews. The data will help the company gain such insights as whether negative reviews disclose if members threaten to delete the app or whether positive reviews contain discussion of fairness as it relates to bot activity.

Thus, the goal of this capstone project is to generate insights from the SNKRS app reviews and build predictive models that deliver valuable insights to Nike stakeholders who are looking for quantifiable ways to increase engagement with their members.

Based on the findings of this project, Nike would be better equipped to address the most frequently discussed topics featured in the SNKRS app reviews, particularly during times surrounding high-profile launches. If Nike knew which topics predict a positive or negative review, the brand could produce more original content that speak to those successes and failures.

This capstone could be expanded into a second project in which a recommendation system is built that analyzes the key words contained in an app user's review. Based on those key words, the system could then offer the reviewer a short list of topic choices that best represent their review. This would provide an opportunity for other users to quickly filter through the 479K reviews to find those that most align to addressing their individual needs.

Data

The data that will be used for this project contains all US reviews submitted to the official SNRKS app (iOS and Android versions) until January 2019, equating to about 10K reviews in total. The dataset variables include:

- the date of the review,
- the numerical rating,
- the text of the review,
- the review's title,
- the reviewer's username,

- and version of the app in place when the review was submitted.

The dataset was scraped from the official Nike SNKRS app (iOS and Android versions) using the Beautiful Soup Python library to parse and extract data from the app reviews displayed via the Sensor Tower website and the Selenium package was used to automate the web browser interaction from within a Python script.

Approach to Solving the Problem

Does the quantity of the reviews change surrounding the release of a high-profile launch?

To measure the changes in quantity of remarks over time – not just their occurrence – we can tell a richer story about the possible level of interest surrounding a high-profile launch, using exploratory text analysis to filter the data based on dates that surround popular launches.

What topics/themes/trends occur most frequently across all reviews?

To provide deeper insights, a topic modeling script will be run that discovers topic scenarios embedded within the text data. Topic modeling is an NLP-based statistical model used for discovering the abstract "topics" that occur in a collection of documents. This approach groups text data by their main themes and involves the use of probability based on word frequencies.

*Which words ("fair" or "bot") embedded in the reviews **predict a positive or negative rating**?*

Using a lexical dispersion technique, we will measure a single word's importance by revealing its location across all reviews. This positional information can be displayed using a dispersion plot, to help visualize how many words from the beginning of a sentence it appears. Because people tend to absorb only the first three words of a review, this will show us which words are valuable to include in a random forest classifier model that predicts the reviews' sentiment.

*Which topics/themes/trends **predict a positive review and which predict a negative review**?*

After utilizing a topic modeling script to discover the main topics embedded within the reviews contained in the dataset, we will input the topics into a random forest classifier model that predicts the reviews' sentiment.

*Does the mention of "deleting the app" or "not entering again" **predict a negative review**?*

Using the ngrams function of the textblob library in Python, we can quickly extract trigrams or three-word phrases such as "deleting the app" or "not entering again" from our reviews and include these trigrams in a random forest classifier model that predicts the reviews' sentiment.

Deliverables

The deliverables for this project will include code, paper documentation, and a visual slide deck presentation – all of which will be posted on my **GitHub repository** and **LinkedIn profile**.