



INSTITUTE FOR APPLIED  
COMPUTATIONAL SCIENCE  
AT HARVARD UNIVERSITY

# Product Review Analysis With pattern

Capstone Project

Team: Sehaj Chawla, Taro Spirit, Lotus Xia, Heather Liu



# Motivation



is an ecommerce accelerator. It helps businesses grow faster and sell globally on ecommerce marketplaces.

How? For example, help them make better decisions through AI-supported insights and reporting.

**Reviews are important!** They give an insight into customer preferences, suggestions, complaints, which ultimately can help businesses

- 1) predict how well their products will do in the future,
- 2) adapt their marketing strategy or even their products.



# Motivation



is an ecommerce accelerator. It helps businesses grow faster and sell globally on ecommerce marketplaces.

How? For example, help them make better decisions through AI-supported insights and reporting.

**Reviews are important!** They give an insight into customer preferences, suggestions, complaints, which ultimately can help businesses

- 1) predict how well their products will do in the future,
- 2) adapt their marketing strategy or even their products.



## Motivation



is an ecommerce accelerator. It helps businesses grow faster and sell globally on ecommerce marketplaces.

How? For example, help them make better decisions through AI-supported insights and reporting.

**Reviews are important!** They give an insight into customer preferences, suggestions, complaints, which ultimately can help businesses

- 1) predict how well their products will do in the future,
- 2) adapt their marketing strategy or even their products.



# Problem Statement

- ★ **Predict product sales volumes using review data**
  - Review metadata: review ratings, length of texts, number of upvotes on each review, etc.
  - Review texts
- ★ **Extract themes from review texts to gain insights on what keywords or topics are predictive of increasing/decreasing sales volumes**



# Problem Statement

- ★ **Predict product sales volumes using review data**
  - Review metadata: review ratings, length of texts, number of upvotes on each review, etc.
  - Review texts
- ★ **Extract themes from review texts to gain insights on what keywords or topics are predictive of increasing/decreasing sales volumes**



# Problem Statement

- ★ **Predict product sales volumes using review data**
  - Review metadata: review ratings, length of texts, number of upvotes on each review, etc.
  - Review texts
- ★ **Extract themes from review texts to gain insights on what keywords or topics are predictive of increasing/decreasing sales volumes**



# Problem Statement

- ★ **Predict product sales volumes using review data**
  - Review metadata: review ratings, length of texts, number of upvotes on each review, etc.
  - Review texts
- ★ **Extract themes from review texts to gain insights on what keywords or topics are predictive of increasing/decreasing sales volumes**





# Data Description

★ Best seller rank history data

★ Review data

- ~ 9991 Products on Amazon
- Vitamins and Dietary Supplements category
- Time spread: 2017-07 ~ 2021-07



# Data Description

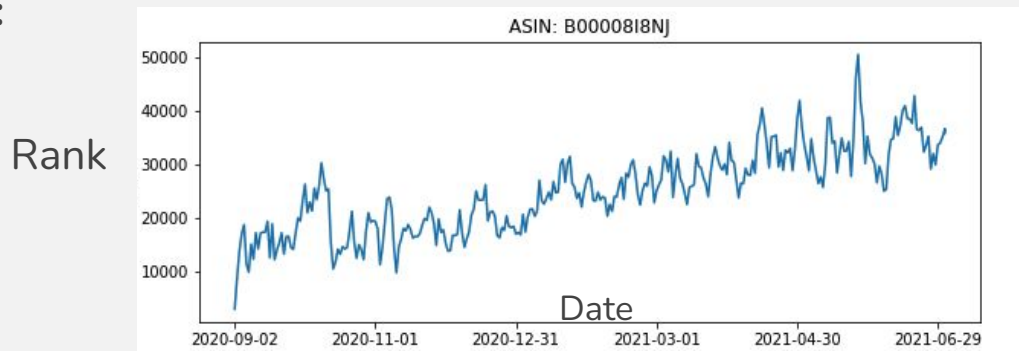
- ★ Best seller rank history data
- ★ Review data
  - ~ 9991 Products on Amazon
  - Vitamins and Dietary Supplements category
  - Time spread: 2017-07 ~ 2021-07

# Data Description

## ★ Best seller rank history data

- Historical data of best seller rank observations for Amazon products
- Date, Average price over 180 days, Best seller rank, etc.

E.g.:





# Data Description

## ★ Review data

- A collection of amazon reviews for products
- Review metadata: date, review ratings, length of texts, number of upvotes on each review, etc.
- Review texts

**E.g.:**

*“Horrible product, my mother in law ended up in the hospital with a severe allergic reaction. She had to be in the ICU for a couple of days. Please be careful with this product.”*

*“I bought this for my father , He swears by it that it helps his joints feel better. His neighbor has been using it for a year now and. No more aches & pains also.”*

*“Pills are not correct as in the picture shown.”*



# Models

- ★ **Task 1: Predict product sales volumes using review data**
  - Model 1: Predicting sales without NLP using product and review metadata
  - Model 2: Using NLP in an independent regression model with no time series element
  - Model 3: Combining our regression model with LSTM structure to add time series
  
- ★ **Task 2: Extract themes/insights from review texts to determine predictive strength**
  - Unsupervised theme extraction using Latent Dirichlet Allocation (LDA)
  - Using predictive-ness indicators on the created models from Task 1



# Models

- ★ **Task 1: Predict product sales volumes using review data**
  - Model 1: Predicting sales without NLP using product and review metadata
  - Model 2: Using NLP in an independent regression model with no time series element
  - Model 3: Combining our regression model with LSTM structure to add time series
  
- ★ **Task 2: Extract themes/insights from review texts to determine predictive strength**
  - Unsupervised theme extraction using Latent Dirichlet Allocation (LDA)
  - Using predictive-ness indicators on the created models from Task 1



# Models

- ★ **Task 1: Predict product sales volumes using review data**
  - Model 1: Predicting sales without NLP using product and review metadata
  - Model 2: Using NLP in an independent regression model with no time series element
  - Model 3: Combining our regression model with LSTM structure to add time series
  
- ★ **Task 2: Extract themes/insights from review texts to determine predictive strength**
  - Unsupervised theme extraction using Latent Dirichlet Allocation (LDA)
  - Using predictive-ness indicators on the created models from Task 1



# Models

- ★ **Task 1: Predict product sales volumes using review data**
  - Model 1: Predicting sales without NLP using product and review metadata
  - Model 2: Using NLP in an independent regression model with no time series element
  - Model 3: Combining our regression model with LSTM structure to add time series
  
- ★ **Task 2: Extract themes/insights from review texts to determine predictive strength**
  - Unsupervised theme extraction using Latent Dirichlet Allocation (LDA)
  - Using predictive-ness indicators on the created models from Task 1