Office Supplies Recommendation System

Volha Puzikava
October, 2022

Disclaimer

The described analyses fulfill educational purposes only. The hypothetical business case, made-up data and the results of the performed analyses should not be considered as real recommendations of the seller, and have not been approved by any professional organization or trading company.

Overview

Office Supply Products

New school year:

- >50% of consumers shop online
- Record amounts are spent to prepare kids for school

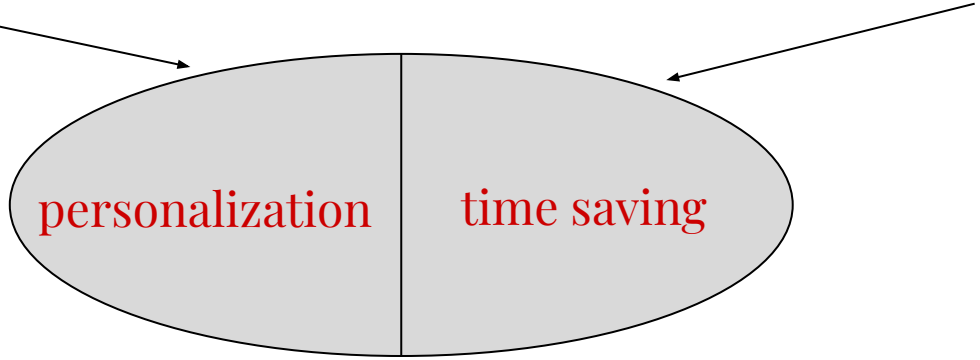
Companies:

- Company with 1-4 employees spends ~ $1,844 per employee yearly
- Well-stocked office = successful office

Overview Cont'd

This project:

1. Recommends office supplies based on the reviews of previously bought products



2. Gives advice, if it is valuable to offer products in the pack of two

Outline

- Business Problem
  - Data
    - Data Preparation and Exploration
      - Modeling
        - Evaluation
          - Conclusions
            - Next Steps

Business Problem

The ***Stationary and Co. Company*** asked:

1. For recommendation system of their office supply products
2. If it is worth to offer products in the pack of two

more sales + high level of personalization

Business Problem Cont'd

Goals:

1. Find the best recommendation model
2. Perform experiment to test the effectiveness of the new package

Data

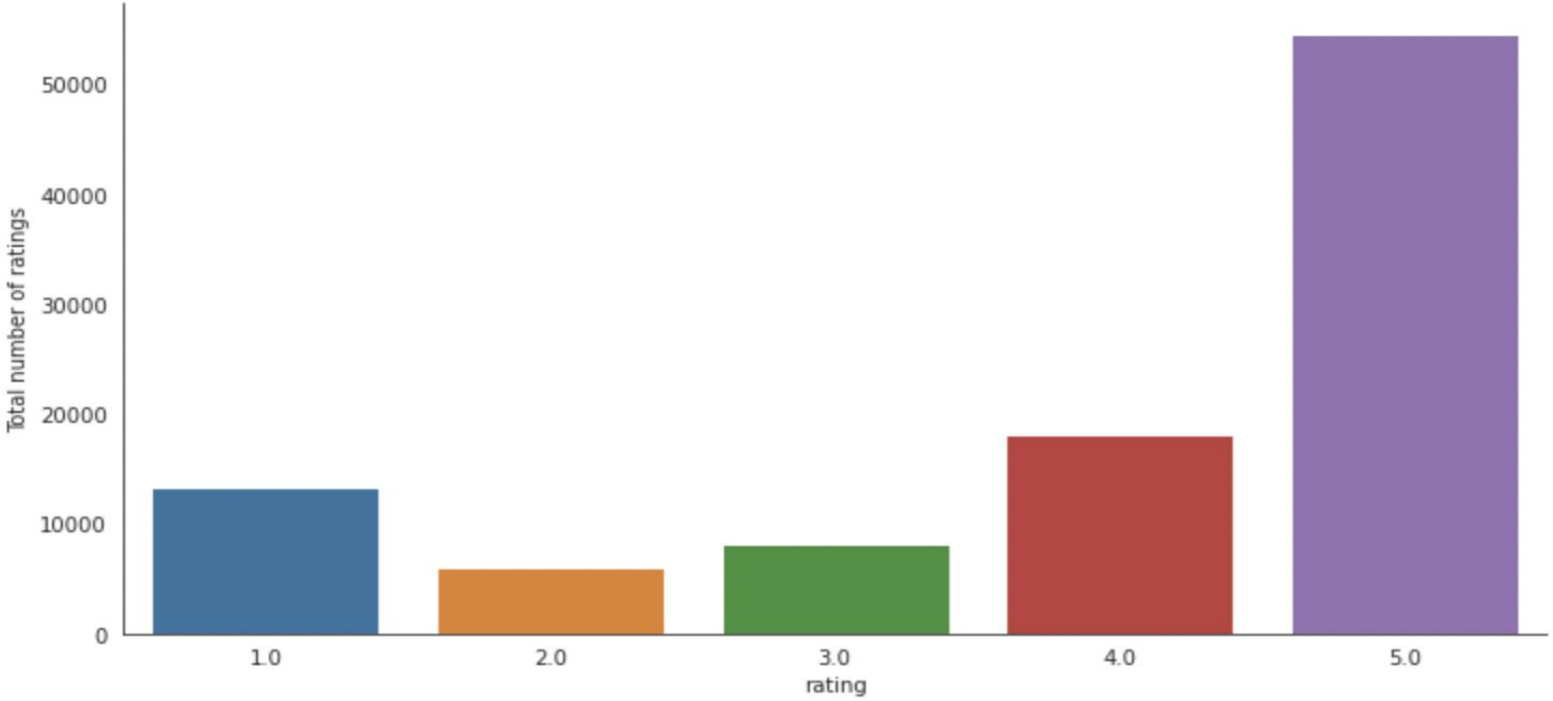
- Taken from Amazon review data (2018) page
- Included 5,581,313 reviews
- Contained meta data for 315,644 office supply products

Data Preparation and Exploration

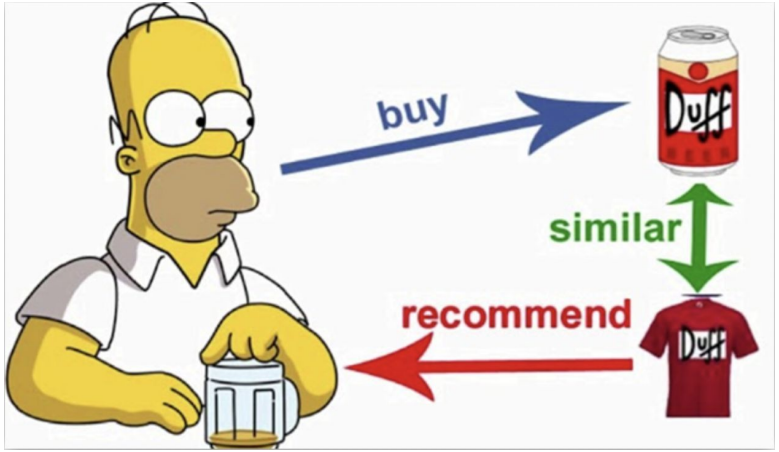
1. Datasets contained:
    - Product_ids,
    - Reviewer_ids,
    - Ratings,
    - Titles,
    - Main categories
2. Datasets merged
3. Random 100,000 rows chosen

Data Preparation and Exploration Cont'd

4. The distribution of ratings plotted



Modeling



1. Models built in
    a. Surprise library,
    b. Spark programming environment

2. *Item-item similarity* performed (search for similar items based on the items users liked)

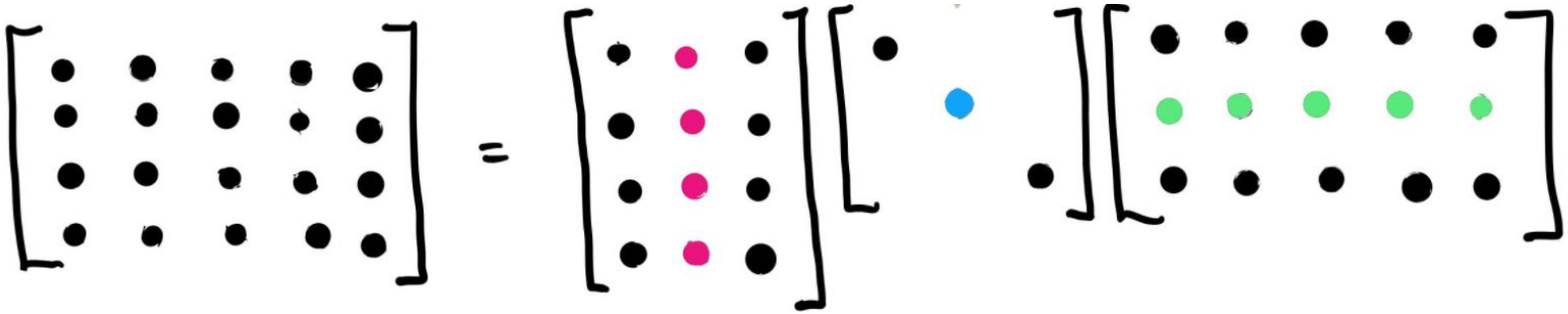
```
Number of users:  94184

Number of items:  27405
```

3. *Different models* tried out

Modeling Cont'd

4. ***Best model - SVD:***
    - Reduced the data into key features (decreased dimensions),
    - Had the *lowest* RMSE (error made while predicting the data)

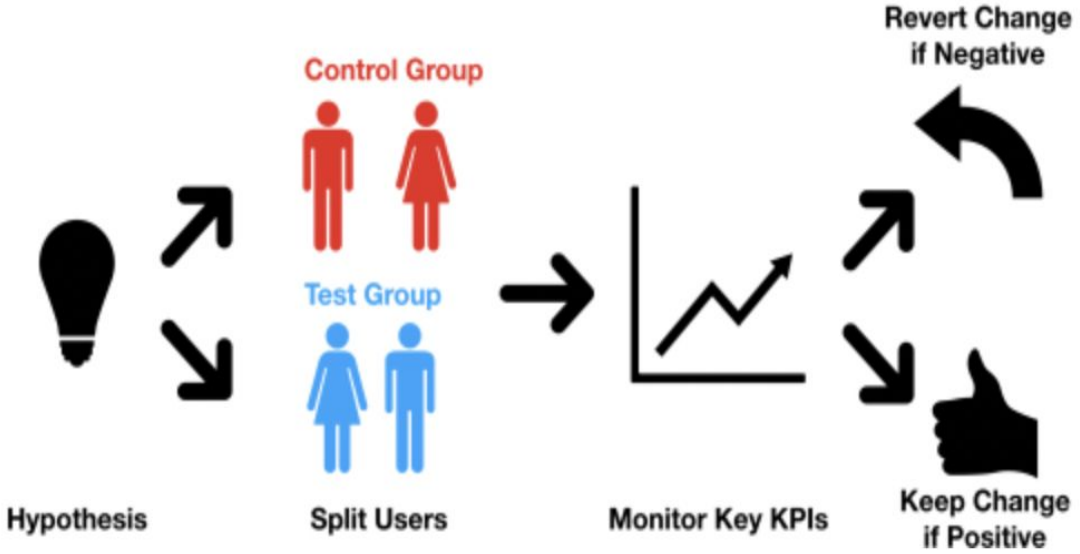


5. *Recommended 10 products* for a user based on reviews

Modeling Cont'd

Experiment

(A/B Testing)



What we know:

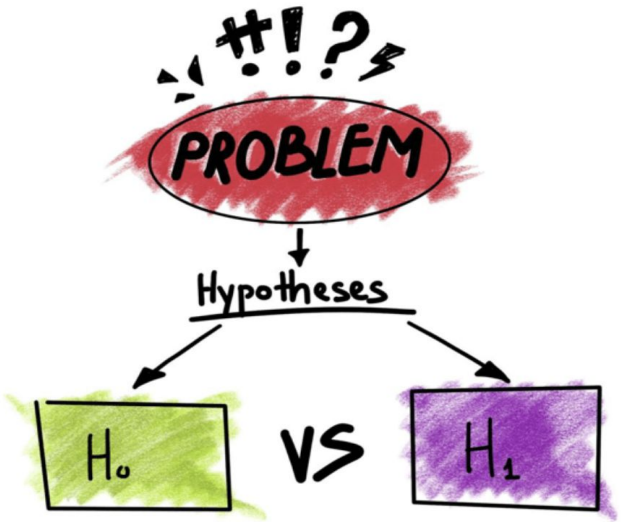
1. Increase in buying rate = 5%
2. ~6.5 million visitors per day
3. ~25% buy products daily

What we do:

1. Two groups: control & experiment
2. Chi-square goodness of fit test

Modeling Cont'd

1. Tested if the probability of success for the new package >= 0.3
2. Min # of customers = 60,279
3. Period of 28 days was analyzed

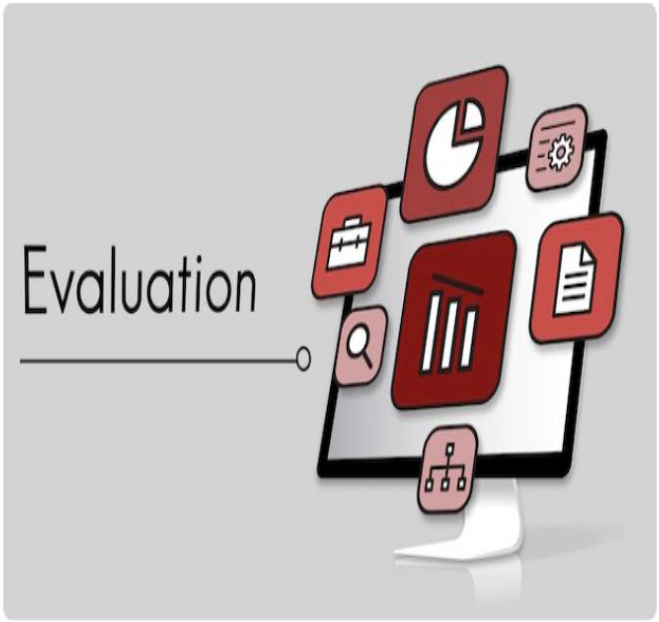


Evaluation

Experiment Results:

a) Pack of two decreased the buying rate by 6.25%

b) Not worth implementing



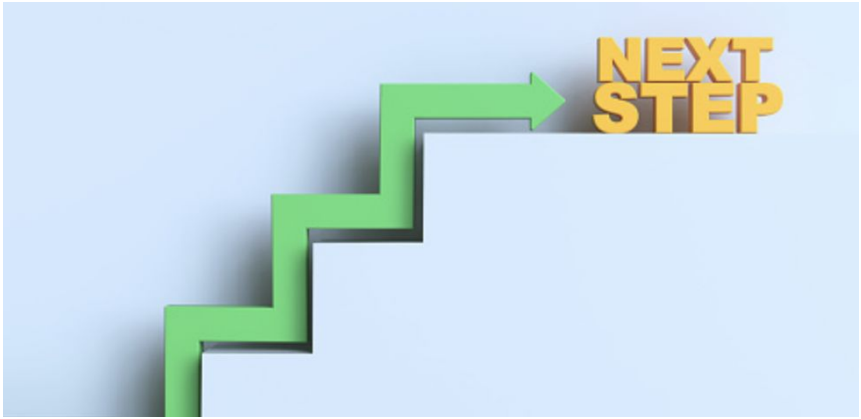
Conclusions

1. Use the SVD model for recommendations

2. Do not implement the new package

Next Steps

1. Tune the models
2. Examples of other ideas to increase sales:
    - User-friendly web site
    - Discounts on customers' birthdays…



Thank You!

Email: helga.mikel@gmail.com

GitHub: @VolhaP87

LinkedIn: https://www.linkedin.com/in/volha-puzikava-23192294a