

HOME TASK ASSIGNMENT REVIEW

AMZ inventory suggestion

Work by Pasichny Denis

1. Data Overview

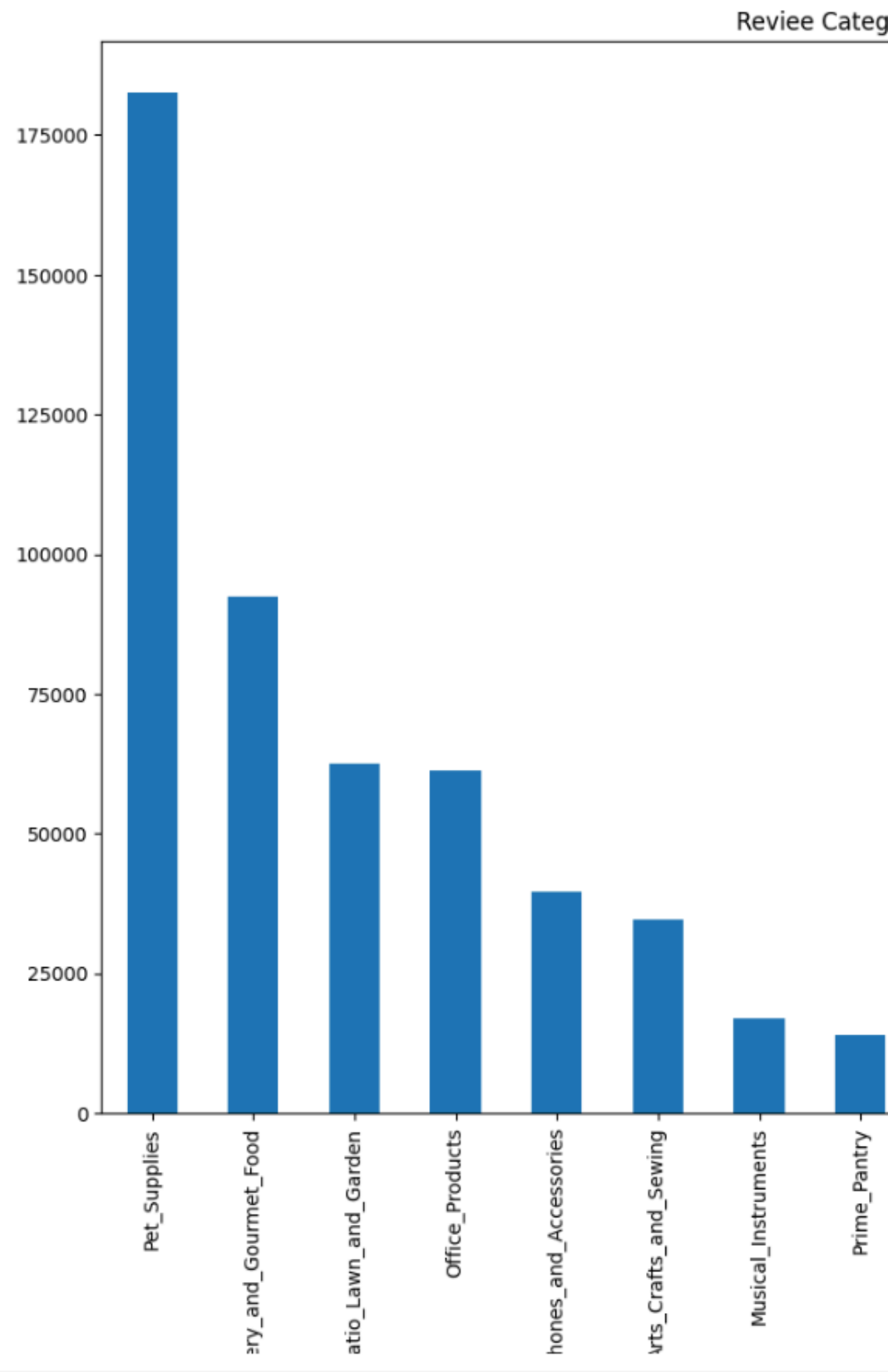
userName	verified	itemName	description	image	brand	feature	category	price	rating	reviewTime	summary	reviewText	vote
Amazon Customer	True	Toblerone Swiss Milk Chocolate Bar, Crunchy Sa...	["Made from deliciously decadent ingredients, ...	["https://images-na.ssl-images-amazon.com/imag...	Toblerone	[]	Prime_Pantry	\$1.63	5.0	2018-01-01	Five Stars	super smooth and yummy with crunchy bits	0
Carla	True	Prince of Peace Organic Tea, Oolong, 100 Tea Bags	['Prince of Peace Enterprises, Inc., founded i...	["https://images-na.ssl-images-amazon.com/imag...	Prince Of Peace	[]	Grocery_and_Gourmet_Food	\$6.40	5.0	2018-01-01	Five Stars	Perfect for kombucha	0
Jaclyn	True	Pet Champion Adjustable No-Pull Harness, Colla...	["The Pet Champion Large/ Extra Large 22-36 in...	["https://images-na.ssl-images-amazon.com/imag...	Pet Champion	['Features Bright Pink Lemonade pattern', 'Dua...	Pet_Supplies	\$7.99	5.0	2018-01-01	she has an odd shape chest and her pull over h...	Finally a harness that fits my puppy. I really...	0
Christinne M	False	Koh-I-Noor Progresso Woodless Colored 24-Penci...	['Koh-I-Noor Progresso Woodless Colored 24-Pen...	["https://images-na.ssl-images-amazon.com/imag...	KOH-I-NOOR	['Rich, pigmented, solid color encased in lacq...	Arts_Crafts_and_Sewing	\$14.18	5.0	2018-01-01	Loving!!	I LOVE THEM!! I bought them at Micheals our of...	0

2. Data Insights

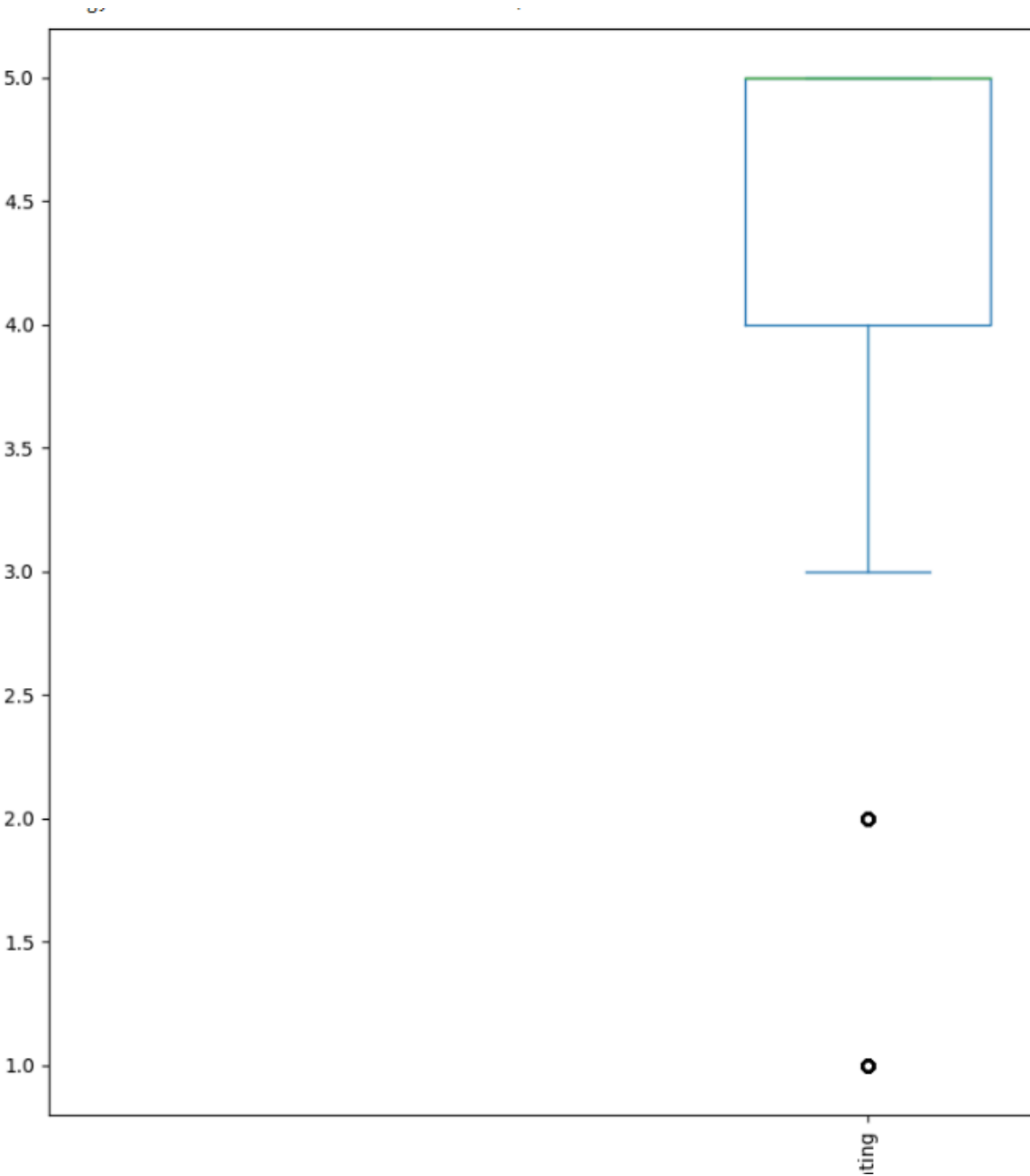
missing values

```
Index: 551159 entries, 0 to 551158
columns (total 14 columns):
Column      Non-Null Count  Dtype
-----
userName    551098 non-null  object
verified    551159 non-null  bool
itemName    551158 non-null  object
description  551151 non-null  object
image       551159 non-null  object
brand       547290 non-null  object
feature     551159 non-null  object
category    551159 non-null  object
price       476825 non-null  object
rating      551159 non-null  float64
reviewTime  551159 non-null  object
summary     550925 non-null  object
reviewText  550764 non-null  object
vote        551159 non-null  int64
```

different sizes of category buckets



vastly positive review ratings



3. Model selection - predicting rating score

We try to **classify** observation into one of five categories : [1,2,3,4,5]



It's a (multi-class)classification problem!

Models: RFC (popular overall and supports multi-class)

SVM (It's a linear model that can be adjusted to multi-class)

Home task Error : Didn't convert label to categorical (stayed numerical)

From this point on in the task, one of two things happened:

1. I didn't(partially) solve this part of the task

OR

2. I don't have output to present anymore

4. Comparison methodologies

classification problem  common comparison metrics: Acc, Precision, recall, F score, AUC

I have calculated ACC, PERCISION, RECALL for each model and put them side by side in a table so it will be easier to compare

```
new_row = pd.Series({'model_name': model.__class__.__name__,  
                    'train_acc':train_acc_score, 'test_acc':test_acc_score,  
                    'test_precis':test_prec, 'test_recall':test_recall})  
res_df = pd.concat([res_df, new_row.to_frame().T], ignore_index=True)
```

Note: During task execution, SVM didn't finish running, therefore no comparison table is available

Home work error: Did not explicitly calculating metrics for each class on it's own

5. Service deployment - Flask app

User decides to enter his observation



The best performing model(e.g SVM) is called to predict 'rating'



observation and it's new rating are forwarded to a recommendation system



best 5 fitting items from inventory are presented to the User



Lastly, update database and item-user matrix

Basic recommendation system for home task

1. Compute user-item interaction matrix (rows: total users, cols: all items in db w/o duplicates, values: ratings)
2. Rate item-item similarity score based on users preferences
3. Display top 5 rated items for our user