Customer Purchase Journey Prediction

- **Aditya Sumbaraju**
  DSC 680 Applied Data Science
  Dr. Brett Werner
  Jan 08, 2021

Table of Contents

Topic

Customer Purchase Journey Prediction

Background

In recent years analyzing shopping baskets has become quite appealing to retailers. Advanced technology made it possible to gather information on their customers and what they buy. Electronic point-in sales increased the use and application of transactional data in the next basket analysis. Analyzing purchase behavior patterns allows retailers to understand customers' purchase behavior, which helps to adjust promotions, store settings and serve customers better.

About Data

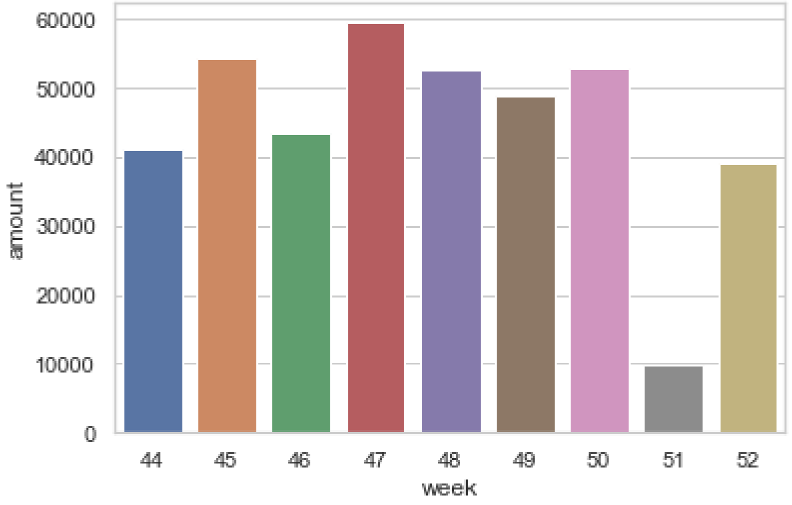
The Ta Feng Grocery Dataset is a supermarket dataset containing
817741 transactions from November 2000 until the end of Feb 2001.
The data set contains information about 119578 shopping baskets that belong to 32266 unique users.
In total, 1129939 items were purchased from available 23812 products.

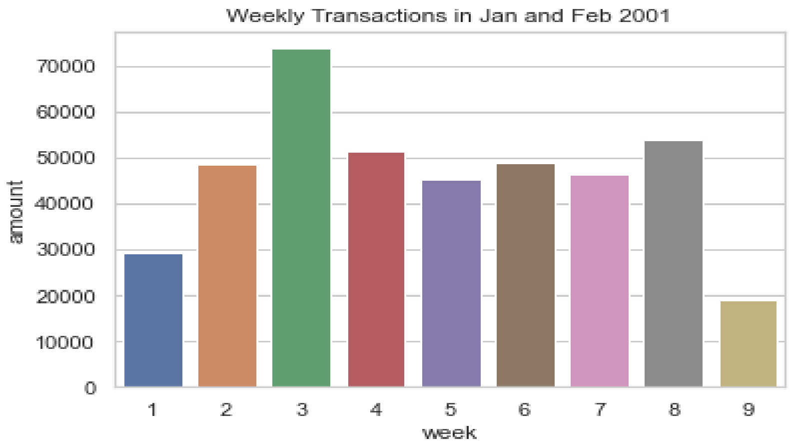
```
]: display (cust_data.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 817741 entries, 0 to 817740
Data columns (total 12 columns):
 #   Column             Non-Null Count   Dtype
---  ------             --------------   -----
 0   transaction_dt     817741 non-null  datetime64[ns]
 1   customer_id        817741 non-null  object
 2   age_group          817741 non-null  object
 3   pin_code           817741 non-null  object
 4   product_subclass   817741 non-null  object
 5   product_id         817741 non-null  object
 6   amount             817741 non-null  object
 7   asset              817741 non-null  object
 8   sales_price        817741 non-null  object
 9   age_label          817741 non-null  object
 10  age_int            817741 non-null  int64
 11  pin_code_int       817741 non-null  int64
dtypes: datetime64[ns](1), int64(2), object(9)
memory usage: 74.9+ MB
```

EDA – Transactions per week

Weekly transactions in Nov and Dec 2000

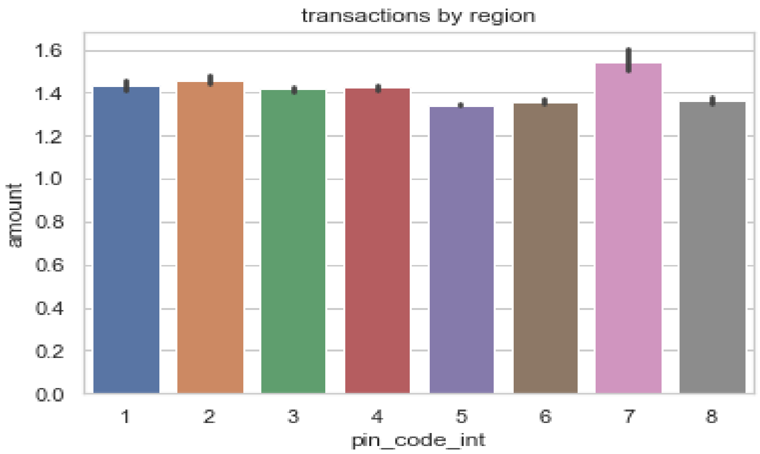


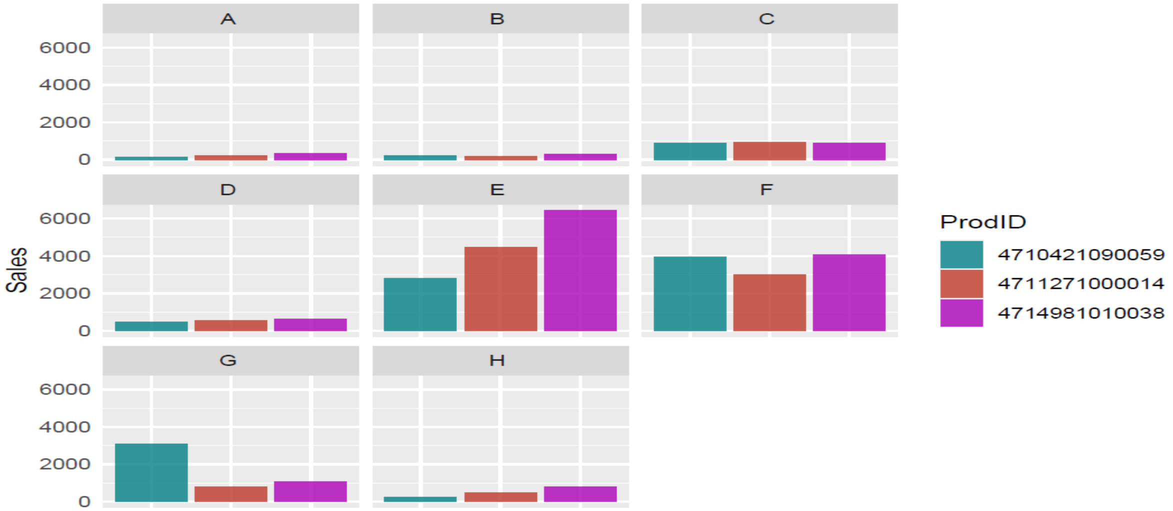


Observation: It is evident that Thanksgiving week accounted for more sales when compared to other weeks. We can see a dip in week 51, and sales look normal in week 52. It is tough to identify the root cause for the drop in week 51 sales from the limited dataset, but we can have a strong assumption by comparing weeks 51 and 52 that inventory refresh could have caused a dip in week 51 sales.

Observation: January 3rd week accounts for more sales than the first week. Ideally, Jan 1st week.2nd and 3rd week are considered peak holiday seasons in the retail domain.

EDA Transactions by region





Observation: Region 7 is contributing more sales followed by 2 and 1. I think we need to apply more marketing strategies in region 5 to boost sales.

Observation:.

1. Region E being the busiest region with the highest number of sales.
2. Regions A and B can be seen to have low numbers of sales. Most Regions show a similar sales profile for all three top products
3. Region G shows the highest number of sales for Product 4710421090059 where as this product is recorded least sales in other regions

Customer Segmentation

Observation: “Summary RFM” represents RFM scores that are easier to read table summarized by the relative RFM score as shown in “Detailed RFM”. The customers with high RFM scores are more important customers to the business. Group customers based on RFM score to classify customers into First Segment, Second Segment, and Third segments for more straightforward interpretation.

- First Segment customers= RFM Score equal or greater than 9
- Second Segment customers = RFM Score between 4 and 9
- Third Segment customers = Anything else

Detailed RFM

customer_id	Recency	Frequency	Monetary	R	F	M	RFM_Score	RMF_Segment
00001069	19	11	1944.0	6	2	3	11	First segment customers
00001113	54	18	2230.0	3	3	3	9	Second segment customers
00001250	19	14	1583.0	6	2	2	10	First segment customers
00001359	87	3	364.0	2	1	1	4	Third segment customers
00001823	36	14	2607.0	5	2	3	10	First segment customers
00002189	57	62	14056.0	3	4	4	11	First segment customers
00003667	21	13	11509.0	6	2	4	12	First segment customers
00004282	47	9	967.0	4	2	2	8	Second segment customers
00004381	103	11	701.0	1	2	1	4	Third segment customers
00004947	81	36	3363.0	2	4	3	9	Second segment customers

Summary RFM

	Recency	Frequency	Monetary	
	mean	mean	mean	count
RMF_Segment				
First segment customers	15.7	37.4	4906.3	18796
Second segment customers	59.9	9.8	1364.5	10647
Third segment customers	97.6	3.6	387.6	2823

Predictive Model and Evaluation

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 simple_rnn (SimpleRNN)      (None, 11)                330

 dense (Dense)               (None, 250)               3000

 dense_1 (Dense)             (None, 11)                2761

=================================================================
Total params: 6,091
Trainable params: 6,091
Non-trainable params: 0
_________________________________________________________________
```

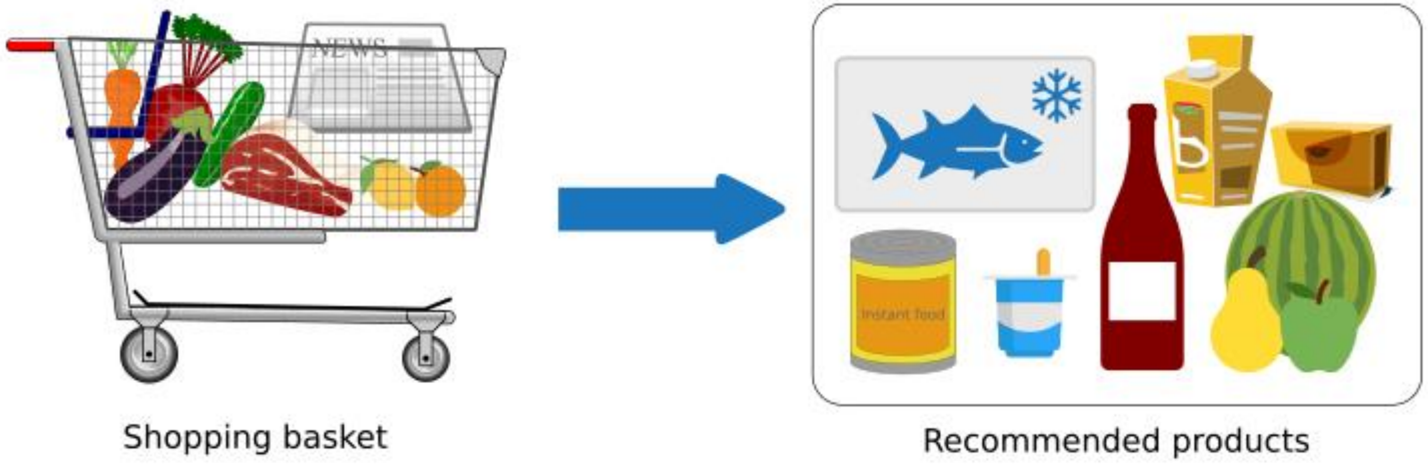
Epoch 995/1000 32/32 [==============================] - 0s 2ms/step - loss: 0.0035 - mae: 0.0203 - acc: 0.7280 - val_loss: 0.0046 - val_mae: 0.0233 - val_acc: 0.6415 Epoch 996/1000 32/32 [==============================] - 0s 2ms/step - loss: 0.0035 - mae: 0.0204 - acc: 0.7189 - val_loss: 0.0046 - val_mae: 0.0223 - val_acc: 0.8432 Epoch 997/1000 32/32 [==============================] - 0s 2ms/step - loss: 0.0035 - mae: 0.0202 - acc: 0.7357 - val_loss: 0.0046 - val_mae: 0.0222 - val_acc: 0.7558 Epoch 998/1000 32/32 [==============================] - 0s 2ms/step - loss: 0.0035 - mae: 0.0202 - acc: 0.7266 - val_loss: 0.0047 - val_mae: 0.0234 - **val_acc: 0.7308** Epoch 999/1000 32/32 [==============================] - 0s 2ms/step - loss: 0.0035 - mae: 0.0209 - **acc: 0.7282** - val_loss: 0.0047 - val_mae: 0.0239 - val_acc: 0.7283 Epoch 1000/1000 32/32 [==============================] - 0s 2ms/step - loss: 0.0035 - mae: 0.0206 - acc: 0.7224 - val_loss: 0.0047 - val_mae: 0.0232 - val_acc: 0.7726

From the above log it is evident that the accuracy is consistently above 70%.the inclusion of all the additional features (Week_number, Amount, Total_sum, Age_group, Pin_code, Unit_price, Log_unit_price) have resulted in significant impact in accuracy.

Model results

	Predicted_cust_id	Week_number	Amount	Total_sum	Age_group	Pin_code	Unit_price	Log_unit_price
0	00020220	1.0	29.0	1444.0	6.0	5.0	1388.0	101.706

Conclusion



This work aims to automate Marketing campaigns using customer segmentation data as an input. The purpose is to simplify the Marketing team's job by avoiding analyzing thousands of rules associating customers and their next item. Recurrent Neural Networks make it easy to find the sequential patterns that return the most probable item sequence. We can retain the customers based on the model-based recommendations by sending appropriate promotional emails. The experiment results depict that RNNs can efficiently predict RFM values of customers. This Model is more likely to be used in recommender systems for next basket recommendation, promotional offers, loyalty programs management, and customer retention programs.

References

- Smartbridge, (2021, January 12). *Market basket analysis 101: Anticipating customer behavior*. Smartbridge. Retrieved December 19, 2021, from https://smartbridge.com/market-basket-analysis-101/
- Kordik, P. (2020, November 14). *Deep learning for recommender systems: Next Basket Prediction and sequential product recommendation*. Medium. Retrieved December 19, 2021, from https://medium.com/recombee-blog/deep-learning-for-recommender-systems-next-basket-prediction-and-sequential-product-recommendation-796228b34dee
- Cavique, L. (2006). *(PDF) next-item discovery in the Market Basket Analysis*. Next-Item Discovery in the Market Basket Analysis. Retrieved December 19, 2021, from https://www.researchgate.net/publication/224693768_Next-Item_Discovery_in_the_Market_Basket_Analysis
- Singh, S. (2021, May 26). *Predicting purchases with Market Basket analysis*. Medium. Retrieved December 19, 2021, from https://medium.com/geekculture/predicting-purchases-with-market-basket-analysis-d6ad2152bf6e
- Guidotti, R., Rossetti, G., Pappalardo, L., Giannotti, F., & Pedreschi, D. (1970, January 1). *[PDF] next basket prediction using recurring sequential patterns: Semantic scholar*. Next Basket Prediction using Recurring Sequential Patterns. Retrieved December 19, 2021, from https://www.semanticscholar.org/paper/Next-Basket-Prediction-using-Recurring-Sequential-Guidotti-Rossetti/b59d6ee45a50a5f5dc7be654cb706897e4ff147c
- H. Salehinejad and S. Rahnamayan, "Customer shopping pattern prediction: A recurrent neural network approach," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), 2016, pp. 1-6, doi: 10.1109/SSCI.2016.7849921.
- Loshin, D. (2013). *Market basket analysis*. Market Basket Analysis - an overview | ScienceDirect Topics. Retrieved January 8, 2022, from https://www.sciencedirect.com/topics/computer-science/market-basket-analysis