BrainStation®

Capstone Project

Prediction of Purchase Time and Quantities by Customer and Product

Cristhian Lima

Dec - 2024

About the Company

Company Details:

The company **‘XYZ’** is located in the USA and **specializes in the manufacturing and printing of foodservice products.**

Its main lines of business include: napkins, cups, bags, and TechnoLiners.

Napkins

Cups

Paper Bags

Techno Liners

Problem Overview:

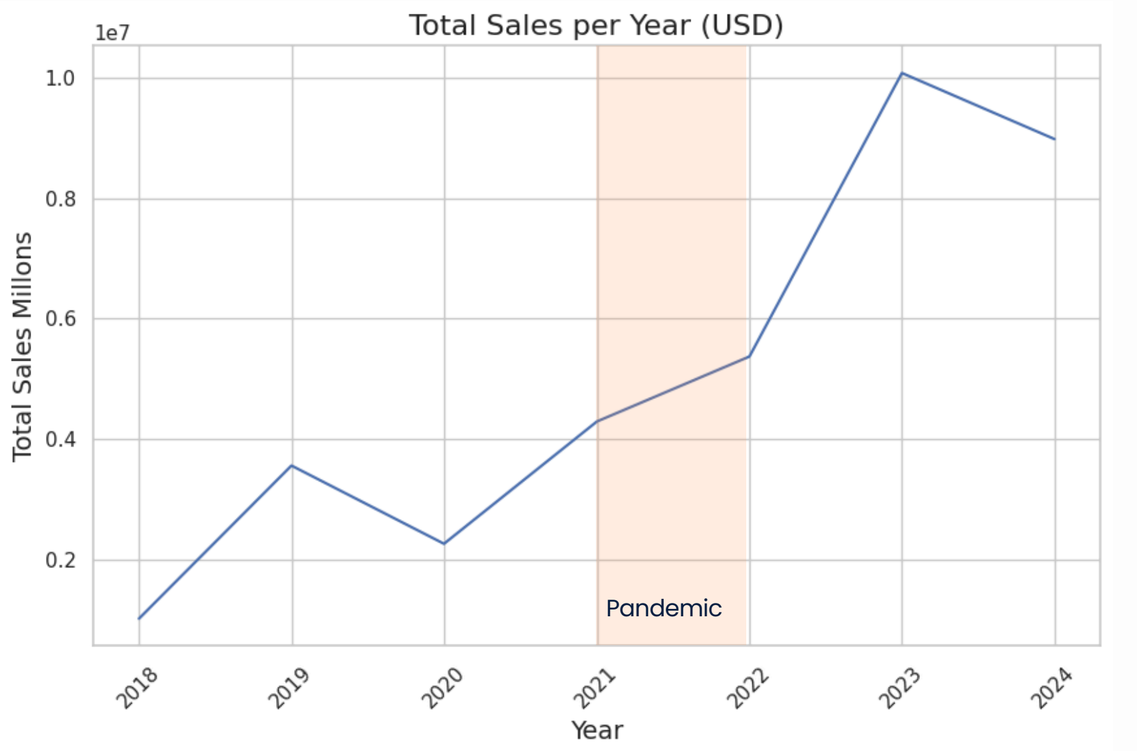
The main challenge is estimating the timing and quantities of the next purchase from key customers.

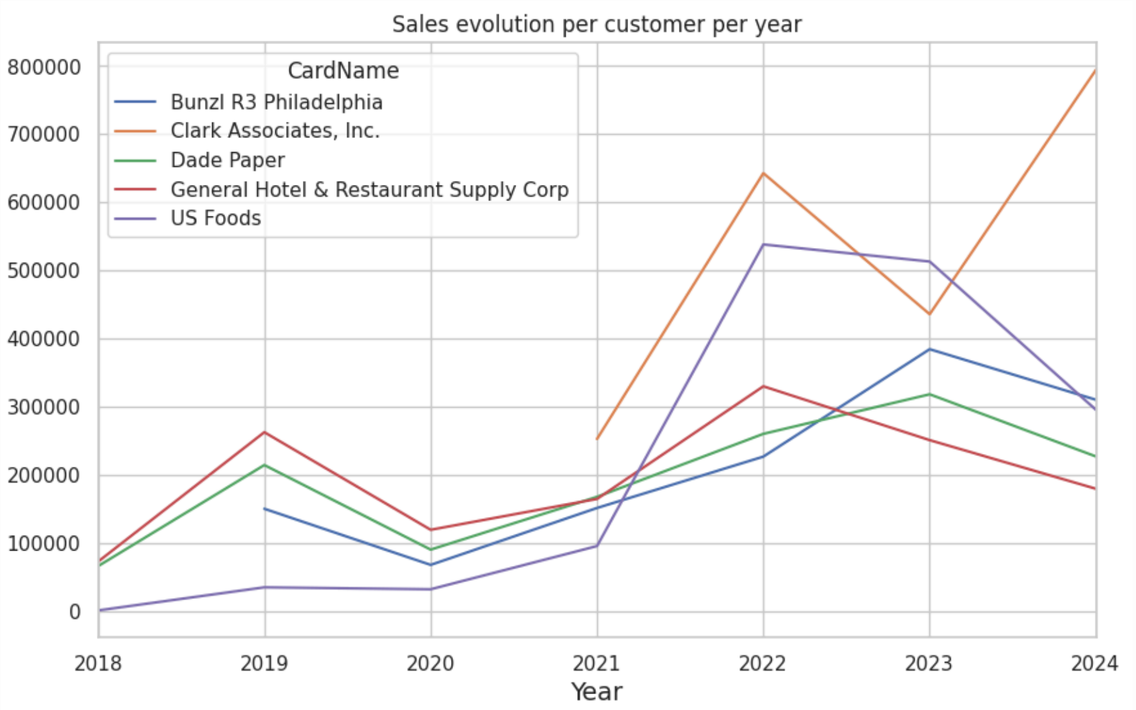
Proposed Solution:

Answer the following two questions:

1. **What is the estimate purchase time for each customer?**
2. **What is the estimate purchase quantity of the items?**

Company Total Sales Over the years





Total Sales: USD 35,571,930

A steady increase in sales evolution, despite pandemic year.

We focus to estimate the purchase time and quantity only for these customers.

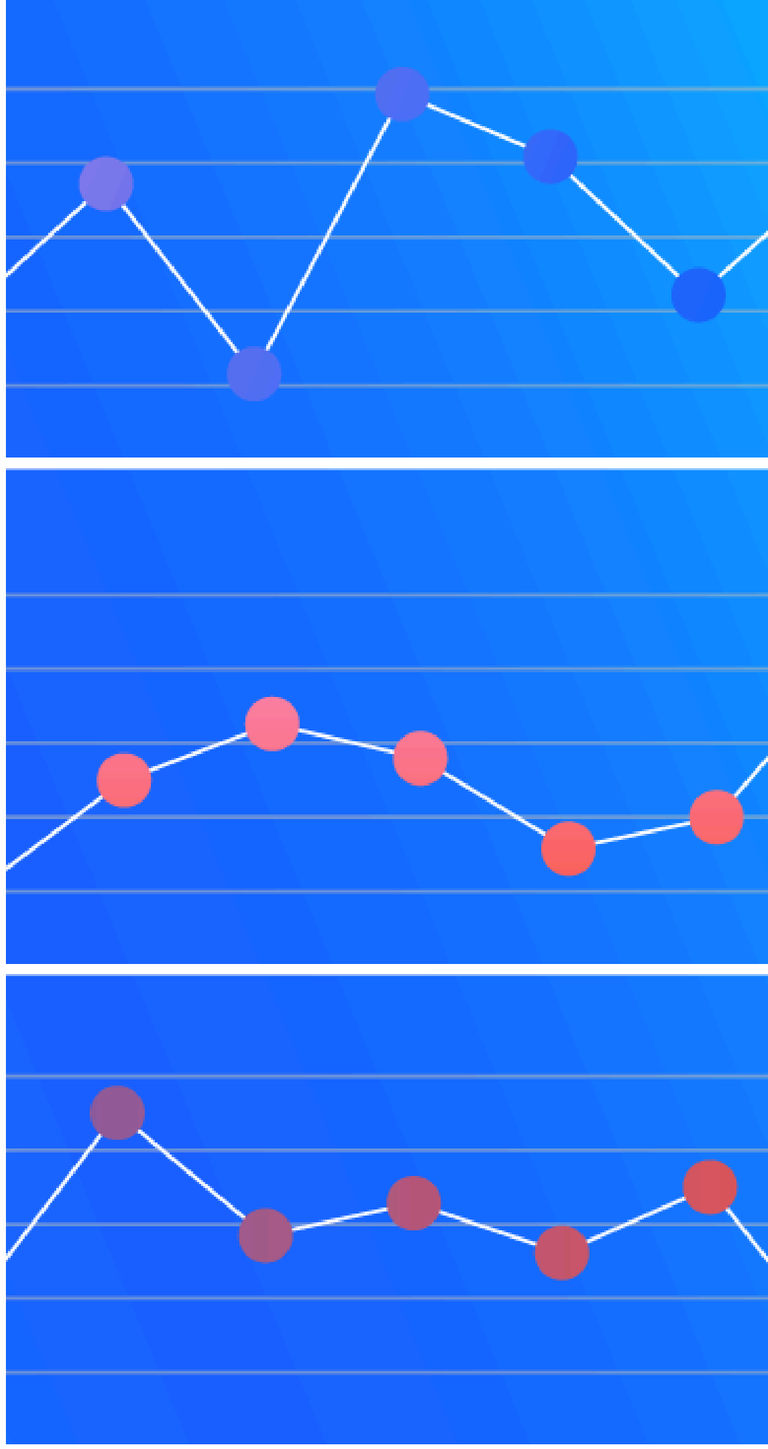
Company Total Sales in Dollars:	35,571,930.29

CardName	DocTotal	Contribution
Clark Associates, Inc.	2,126,590	5.98%
US Foods	1,511,295	4.25%
General Hotel & Restaurant Supply Corp	1,380,795	3.88%
Dade Paper	1,344,886	3.78%
Bunzl R3 Philadelphia	1,291,922	3.63%
	7,655,488	**21.52%**

Modeling Approach

the Approach to calculate the Next Purchase Time and Quantities per Customer

Method	Objective
Simple Average Method	Using the time between every purchase from historical data, calculate the average.
Linear Regression	Based on quantities and prices, calculate the days to the next purchase
Time Series	Calculate quantities to be sold for the next months and identify temporal patterns.



Simple Average and Linear Regression

	DocDate	CardName	ItemName	Real_Dates_Between	AVG_Predicted_Dates	LinReg_Predicted_Dates	AVG_Diff	LinReg_Diff	Better_Method
20	2023-08-30	Bunzl R3 Philadelphia	Dinner Napkin: 2-Ply 15x17, 1/8 Fold. Premium ...	47.0	69	40.398368	-22.0	6.601632	LinReg
21	2023-10-24	Bunzl R3 Philadelphia	Dinner Napkin: 2-Ply 15x17, 1/8 Fold. Premium ...	55.0	69	40.229183	-14.0	14.770817	AVG
22	2023-11-30	Bunzl R3 Philadelphia	Dinner Napkin: 2-Ply 15x17, 1/8 Fold. Premium ...	37.0	69	40.205067	-32.0	-3.205067	LinReg
23	2024-01-31	Bunzl R3 Philadelphia	Dinner Napkin: 2-Ply 15x17, 1/8 Fold. Premium ...	62.0	69	40.658781	-7.0	21.341219	AVG
24	2024-04-25	Bunzl R3 Philadelphia	Dinner Napkin: 2-Ply 15x17, 1/8 Fold. Premium ...	85.0	69	40.658781	16.0	44.341219	AVG
25	2024-05-31	Bunzl R3 Philadelphia	Dinner Napkin: 2-Ply 15x17, 1/8 Fold. Premium ...	36.0	69	39.643674	-33.0	-3.643674	LinReg
26	2024-07-24	Bunzl R3 Philadelphia	Dinner Napkin: 2-Ply 15x17, 1/8 Fold. Premium ...	54.0	69	40.658781	-15.0	13.341219	LinReg
27	2024-08-30	Bunzl R3 Philadelphia	Dinner Napkin: 2-Ply 15x17, 1/8 Fold. Premium ...	37.0	69	40.658781	-32.0	-3.658781	LinReg

Table shows the Average Method and Linear Regression to estimate the Purchasing Time and comparing with Real Values

Better Method

Simple Average: 15 -> is better method

Linear Regression: 13

Quantities Forecast by Time Series

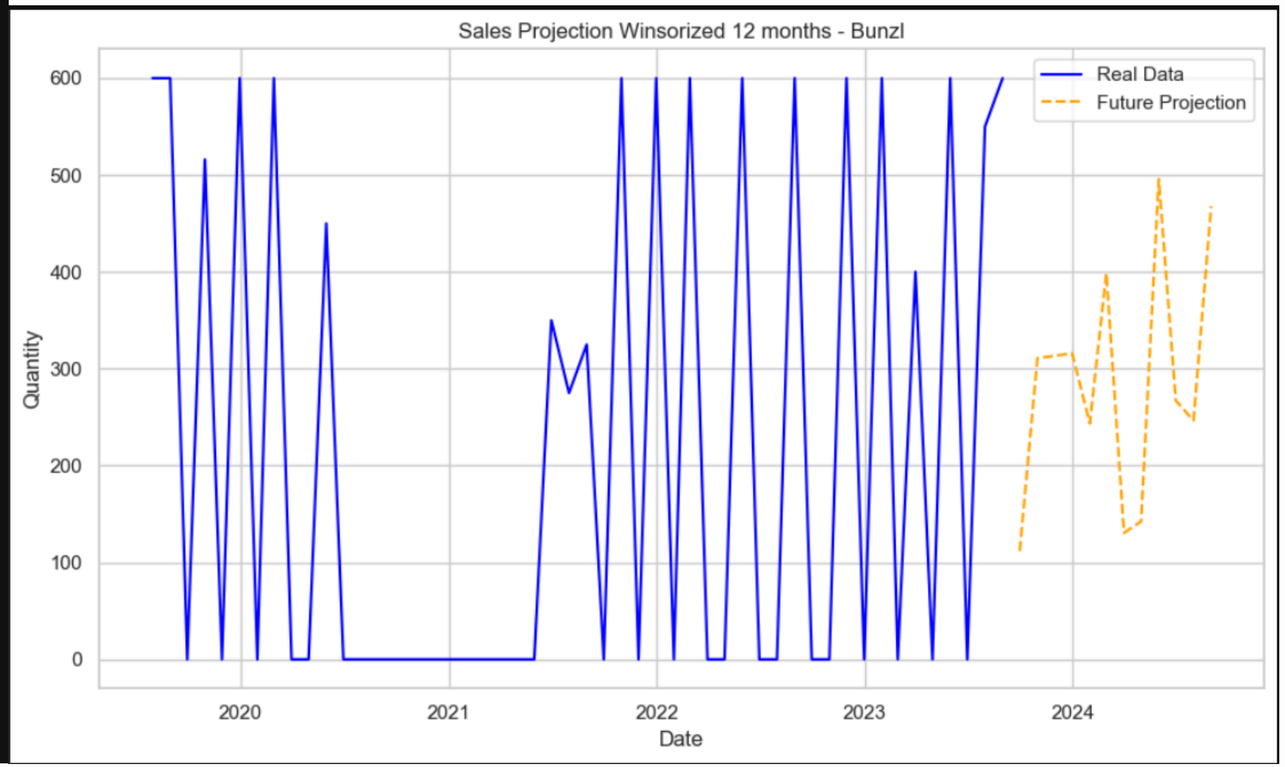
	Year	Month	Quantity
56	2023	9	0.0
57	2023	10	600.0
58	2023	11	600.0
59	2023	12	0.0
60	2024	1	600.0
61	2024	2	0.0
62	2024	3	0.0
63	2024	4	600.0
64	2024	5	600.0
65	2024	6	0.0
66	2024	7	600.0
67	2024	8	600.0

Date	Projection
2023-09-30	111.814970
2023-10-31	310.849377
2023-11-30	313.356006
2023-12-31	315.862635
2024-01-31	243.556765
2024-02-29	398.667560
2024-03-31	130.576967
2024-04-30	142.333596
2024-05-31	495.756892
2024-06-30	267.013521
2024-07-31	246.603484
2024-08-31	467.860113

Right Table: Real Quantities
Left Table: Predicted Quantities by Time Series

The average of the quantities purchased: 350 units.

- The projection by time series is very close to the average.
- Both options, Average and Time Series, deviate significantly from the actual values.
- This projection represent a baseline, the projection model needs to be adjusted.



Expected sales quantities for client Bunzl in 2024

Conclusions

The two primary original questions:

- **Estimated date of the next purchase.**
- **Estimated purchase quantities.**

These models establishes a baseline for data-driven decision-making, enabling the company to better plan operations and proactively address the needs of its key clients.

1. Simple Average Calculation: We calculated the average number of days between each customer's purchases based on historical sales data.
2. Linear Regression: A linear regression model was trained, achieving 43% accuracy in predicting estimated purchase times using quantities and prices as variables.
3. Time Series Analysis: A time series model projected monthly sales quantities for 2024.

Final Recomendations:

- Improve each of the models presented.
- Apply new models and compare them with the current ones.
- Enhance the time series models by applying ARIMA and SARIMA.

THANK YOU