

A Statistical Study on the Sales Trends in the Dairy Goods Industry

Presentation by
Sakshi Mhatre,
Dhruvi Patel,
Ram Rajput.



Introduction

This study focuses on understanding the sales and distribution patterns of different dairy products across various brands and regions.

It aims to analyze how factors like product type, brand, and location influence sales, providing insights into market trends and consumer preferences in the dairy industry.



Index

- Descriptive Statistics
- Visualization of different aspects of data
- Polynomial Regression
- T-Test
- ANOVA
(Analysis of variance)
- chi - Square
- Z - Test
- Conclusion



Categorical analysis

	Location	Farm Size	Product Name	Brand	Storage Condition	Production Date	Expiration Date	Customer Location	Sales Channel
count	4325	4325	4325	4325	4325	4325	4325	4325	4325
unique	15	3	10	11	5	1405	1441	15	3
top	Delhi	Large	Curd	Amul	Refrigerated	2022-01-13	2019-07-27	Delhi	Retail
freq	525	1462	479	1053	2459	9	9	499	1478

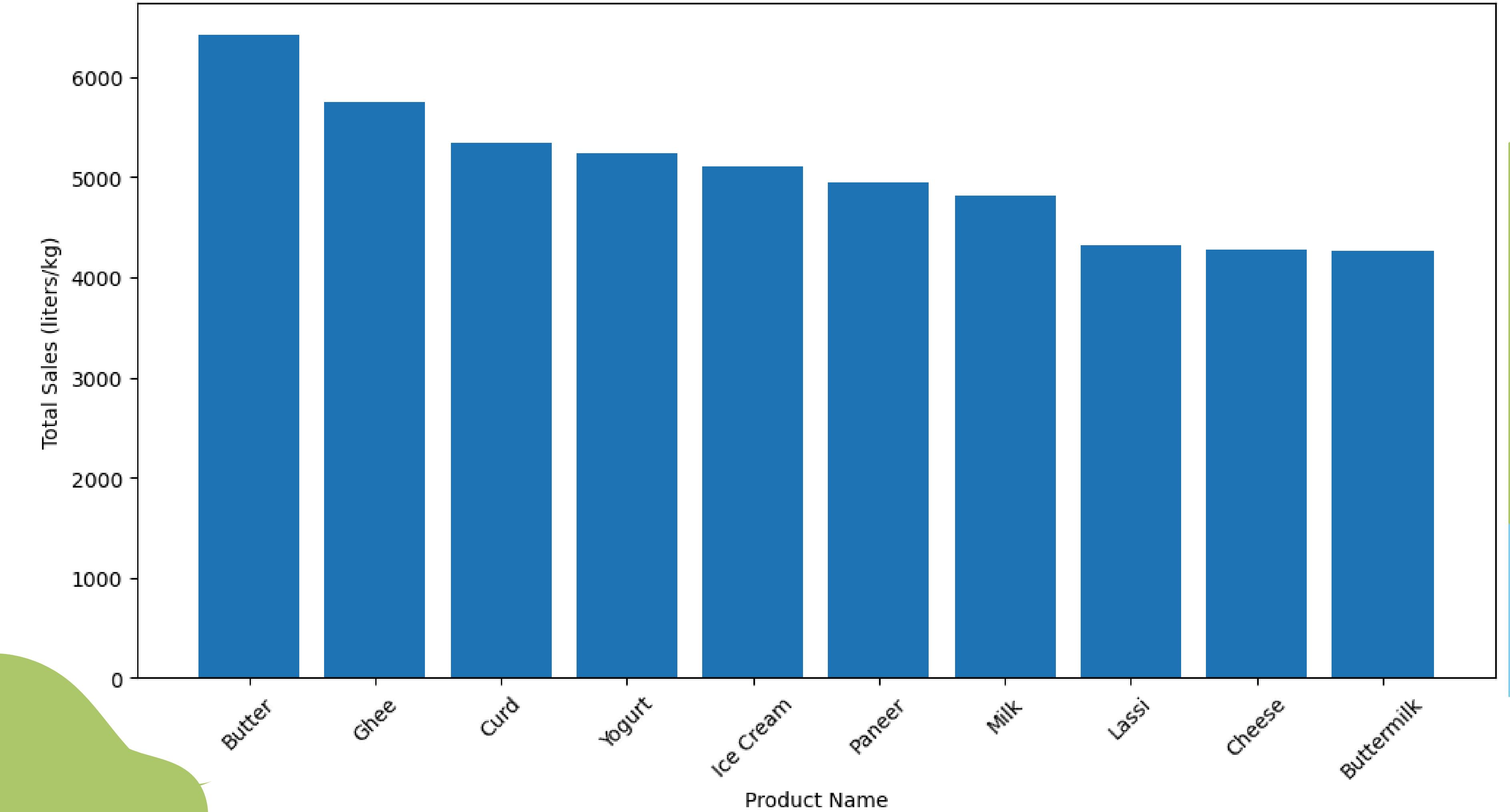


Descriptive Statistics numerical analysis

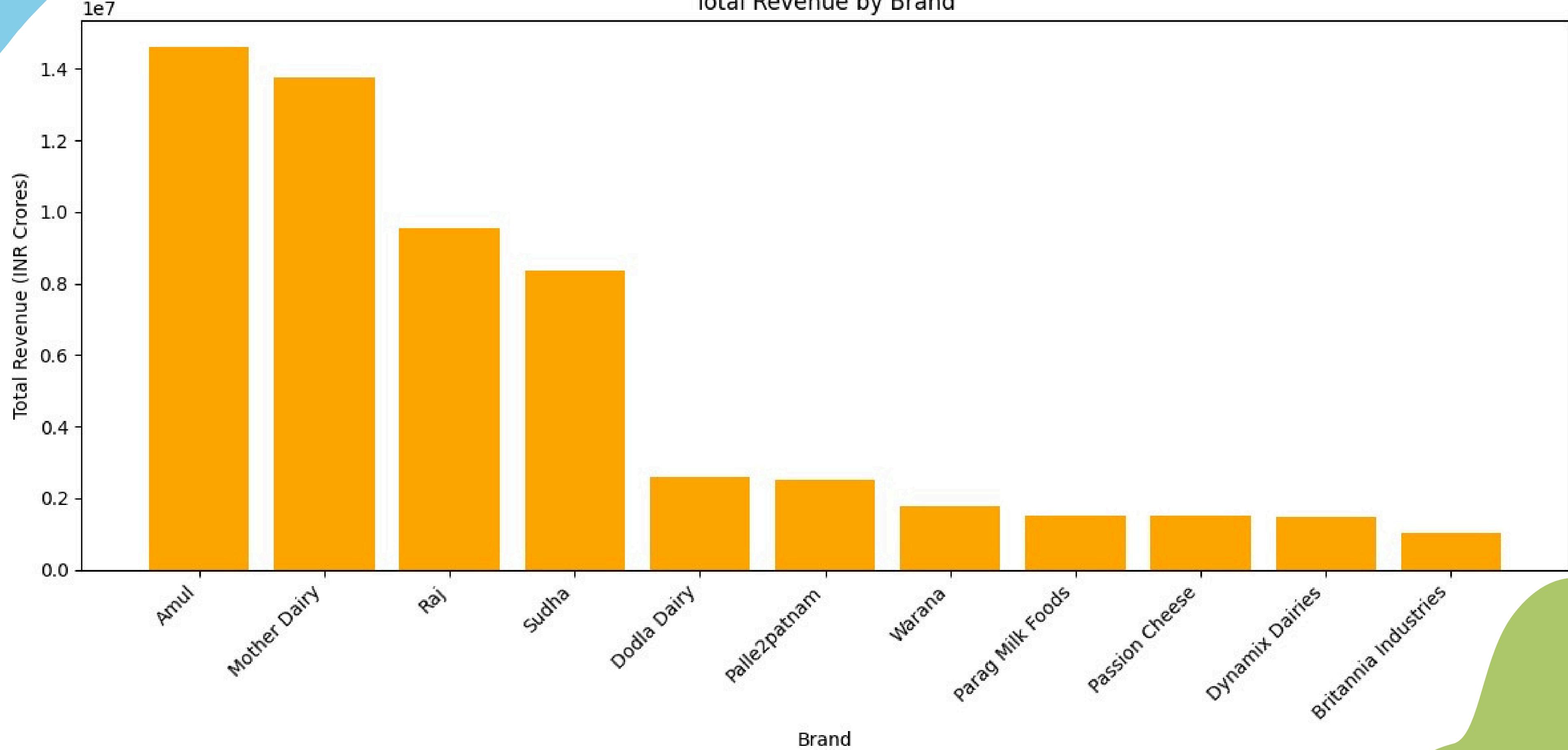
	Total Land Area (acres)	Number of Cows	Product ID	Quantity (liters/kg)	Price per Unit	Total Value	Shelf Life (days)
count	4325.000000	4325.000000	4325.000000	4325.000000	4325.000000	4325.000000	4325.000000
mean	503.483073	54.963699	5.509595	500.652657	54.785938	27357.845411	29.127630
std	285.935061	26.111487	2.842979	288.975915	26.002815	21621.051594	30.272114
min	10.170000	10.000000	1.000000	1.170000	10.030000	42.516500	1.000000
25%	252.950000	32.000000	3.000000	254.170000	32.460000	9946.814500	10.000000
50%	509.170000	55.000000	6.000000	497.550000	54.400000	21869.652900	22.000000
75%	751.250000	77.000000	8.000000	749.780000	77.460000	40954.441000	30.000000
max	999.530000	100.000000	10.000000	999.930000	99.990000	99036.369600	150.000000
	Quantity Sold (liters/kg)	Price per Unit (sold)	Approx. Total Revenue (INR)	Quantity in Stock (liters/kg)	Minimum Stock Threshold (liters/kg)	Reorder Quantity (liters/kg)	
count	4325.000000	4325.000000	4325.000000	4325.000000	4325.000000	4325.000000	
mean	248.095029	54.77914	13580.265401	252.068671	55.826143	109.107820	
std	217.024182	26.19279	14617.009122	223.620870	26.301450	51.501035	
min	1.000000	5.21000	12.540000	0.000000	10.020000	20.020000	
25%	69.000000	32.64000	2916.650000	66.000000	32.910000	64.280000	
50%	189.000000	54.14000	8394.540000	191.000000	56.460000	108.340000	
75%	374.000000	77.46000	19504.550000	387.000000	79.010000	153.390000	
max	960.000000	104.51000	89108.900000	976.000000	99.990000	199.950000	

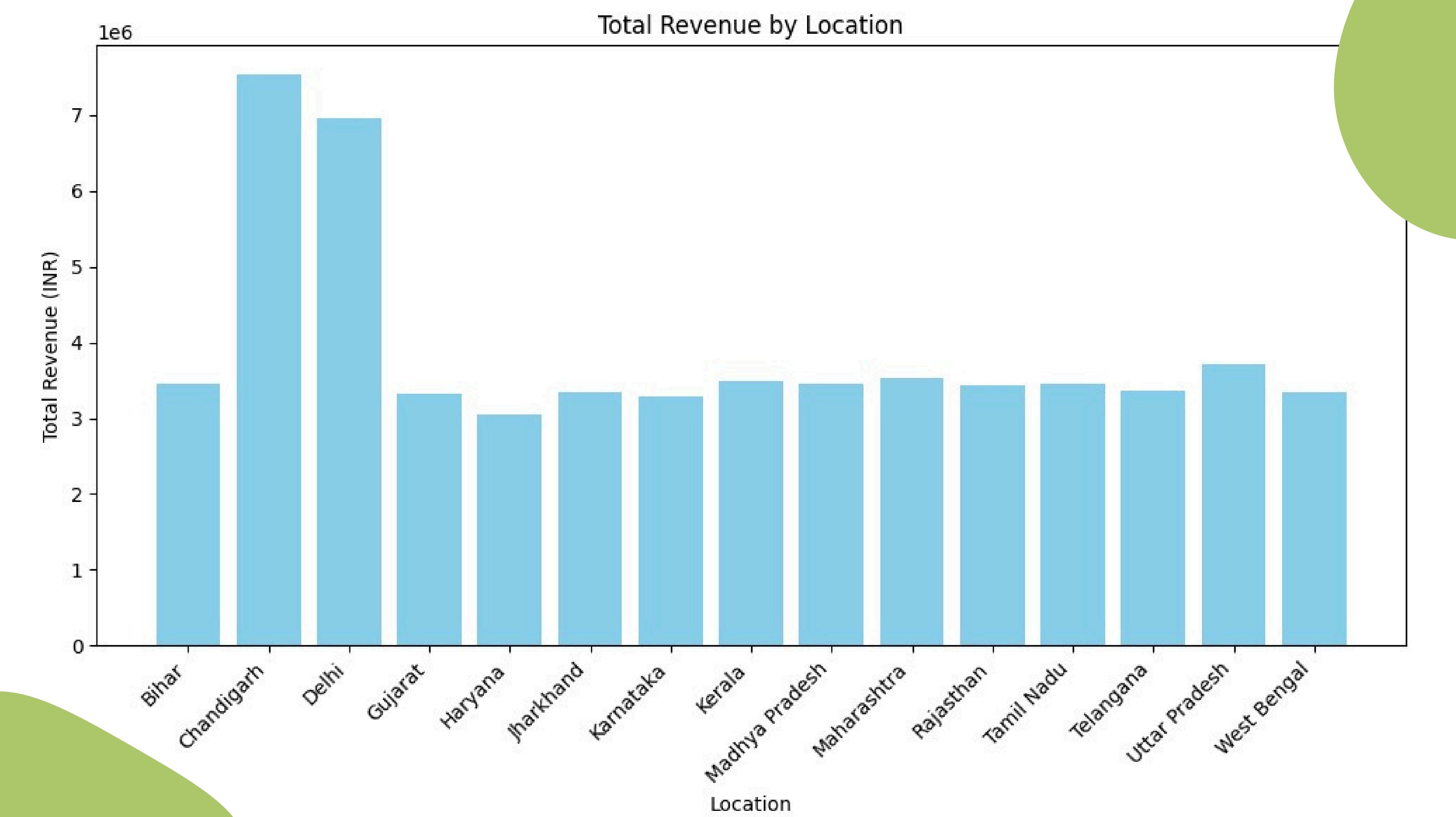


Total Sales of Dairy Products by Product Name



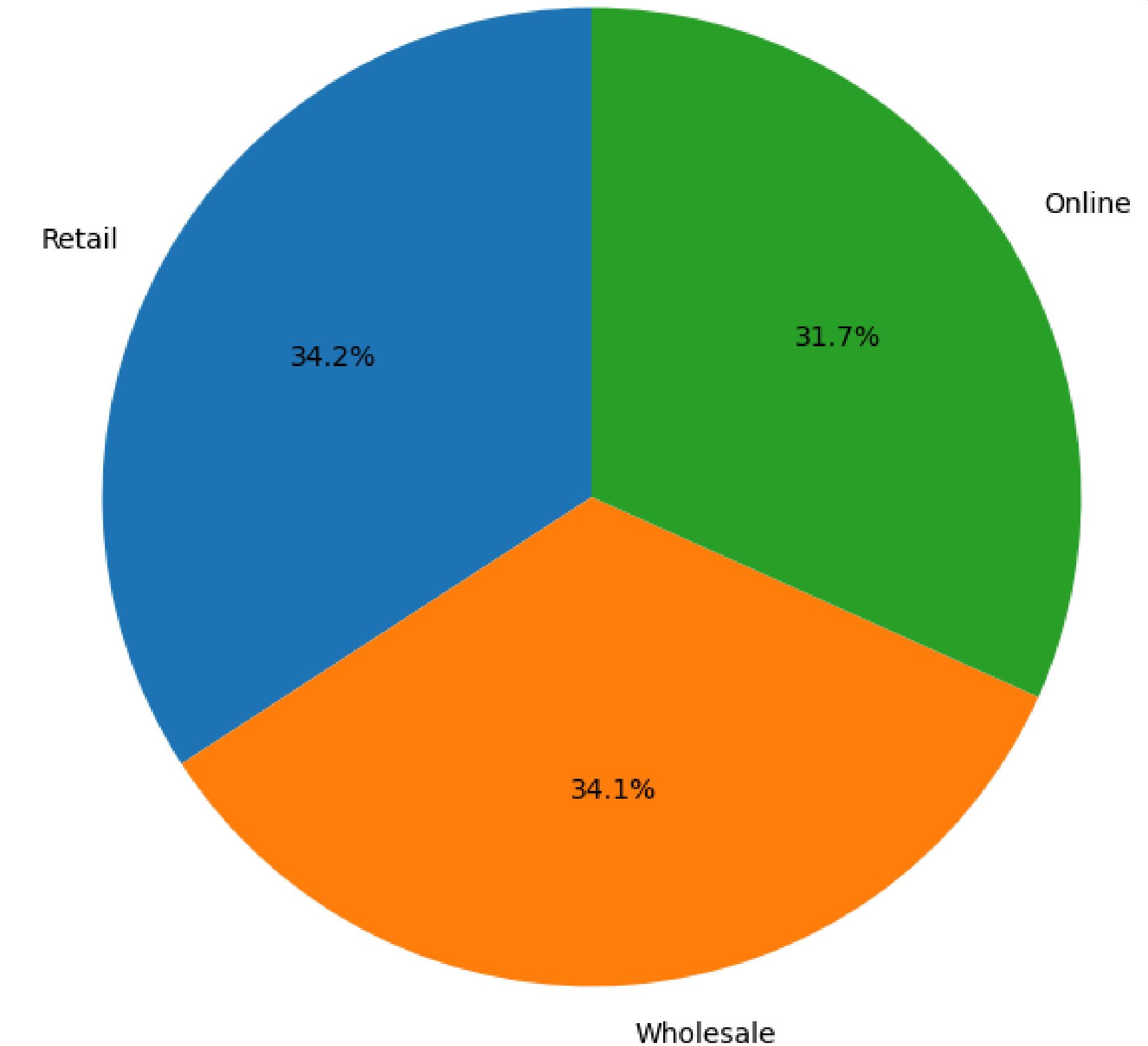
Total Revenue by Brand

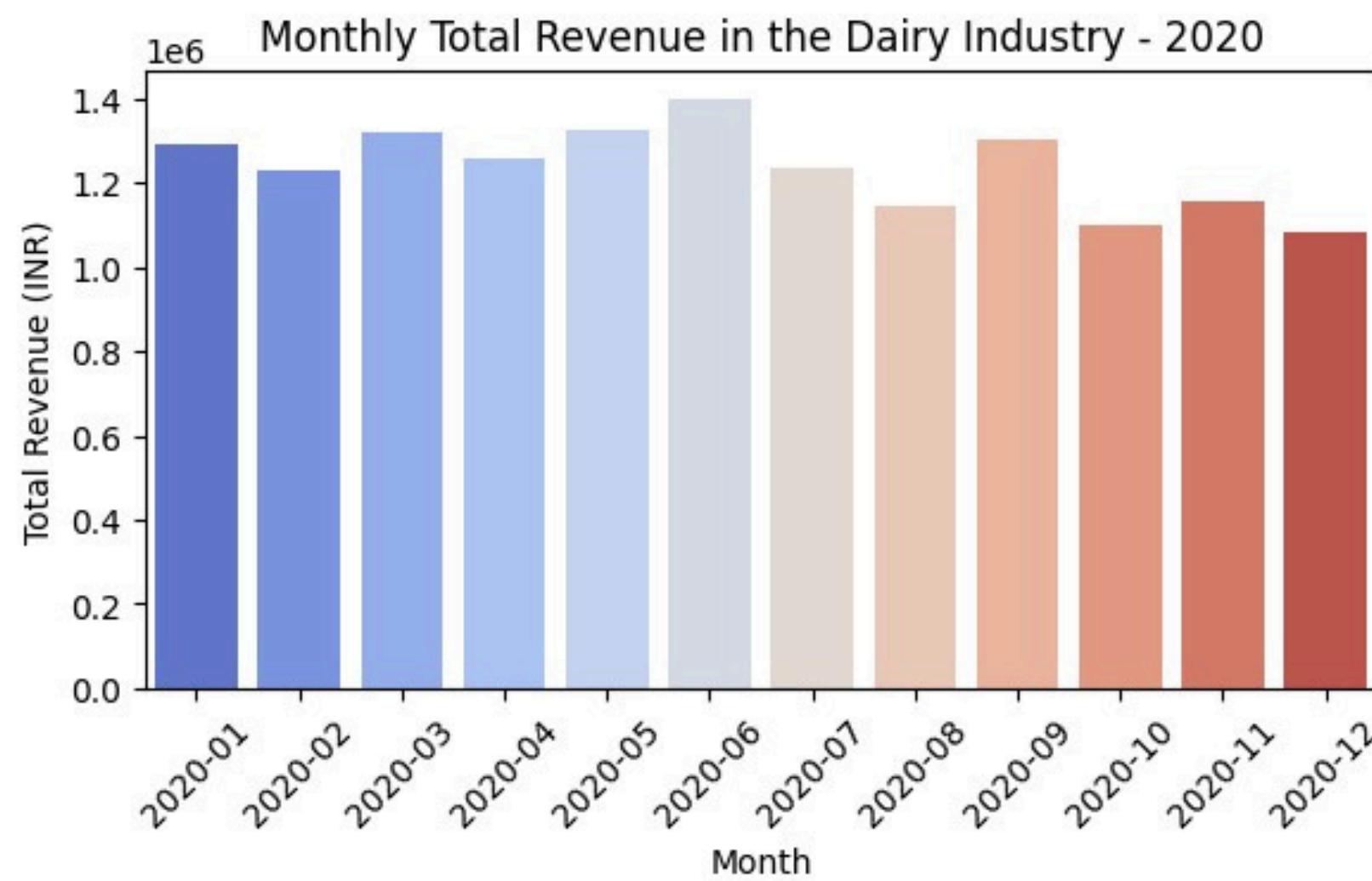
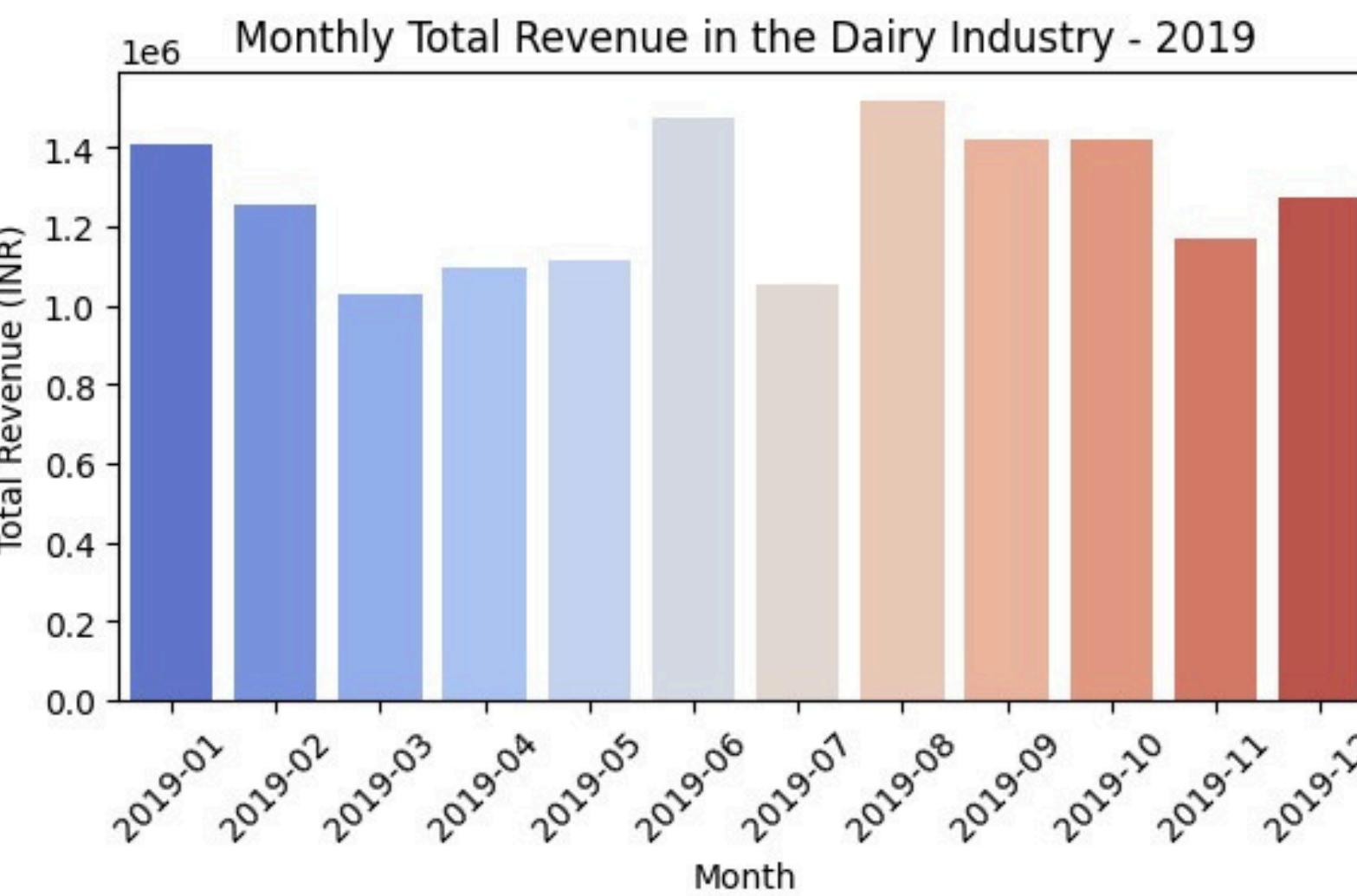
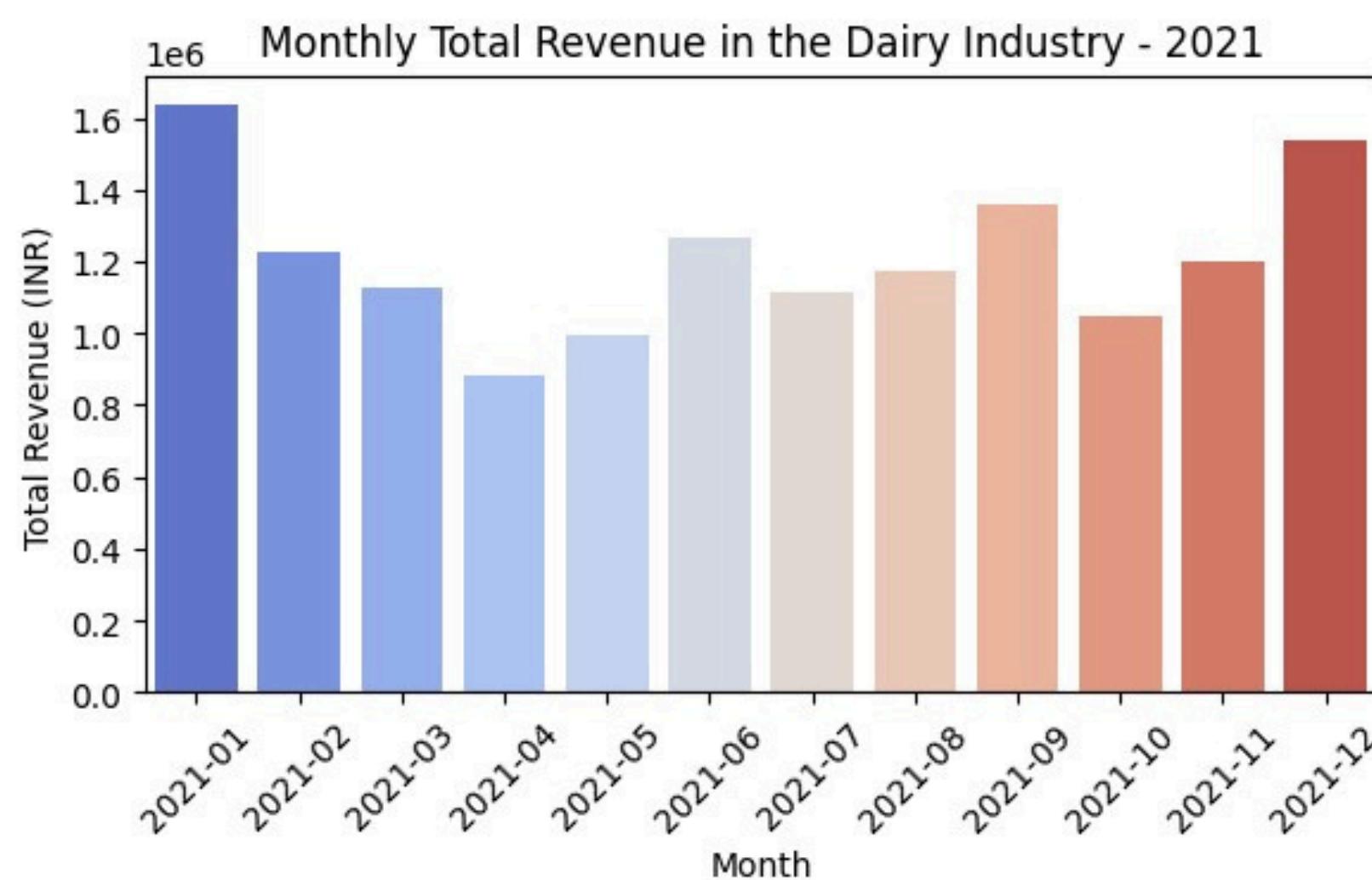
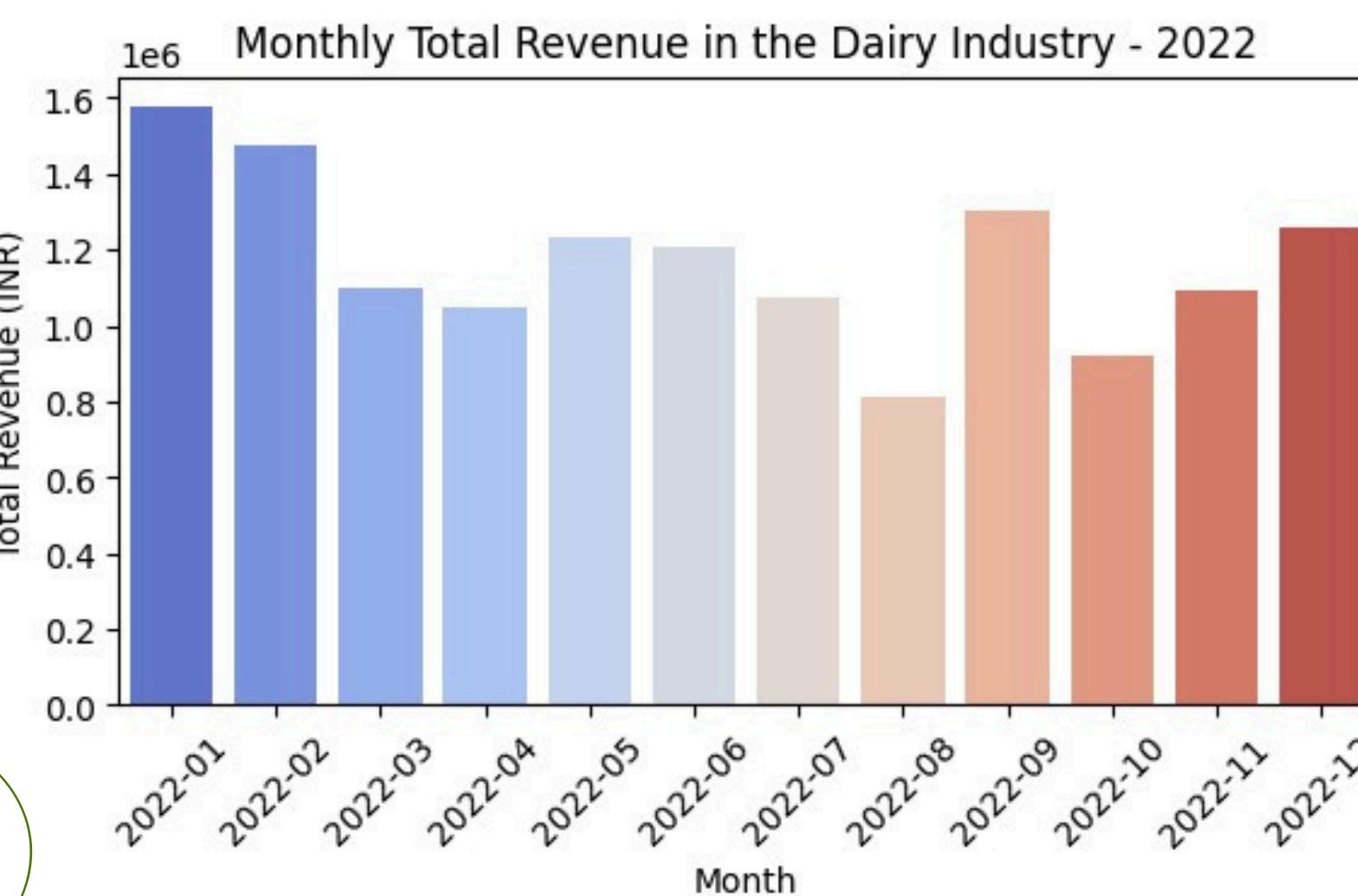




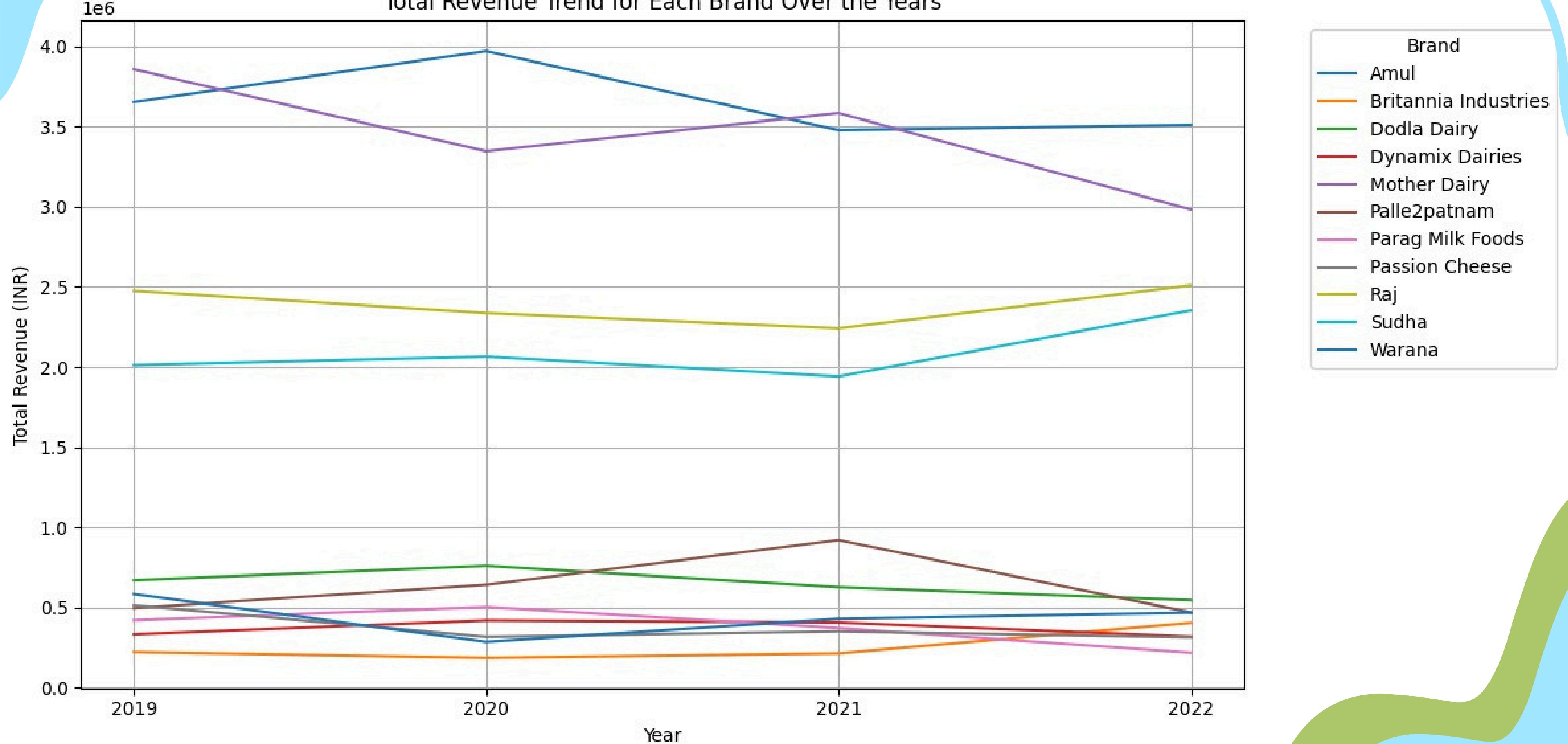
Distribution of Sales Channels

Most items sold via
retail(34.2%),
Online(31.7%)
and
wholesale (34.1%).

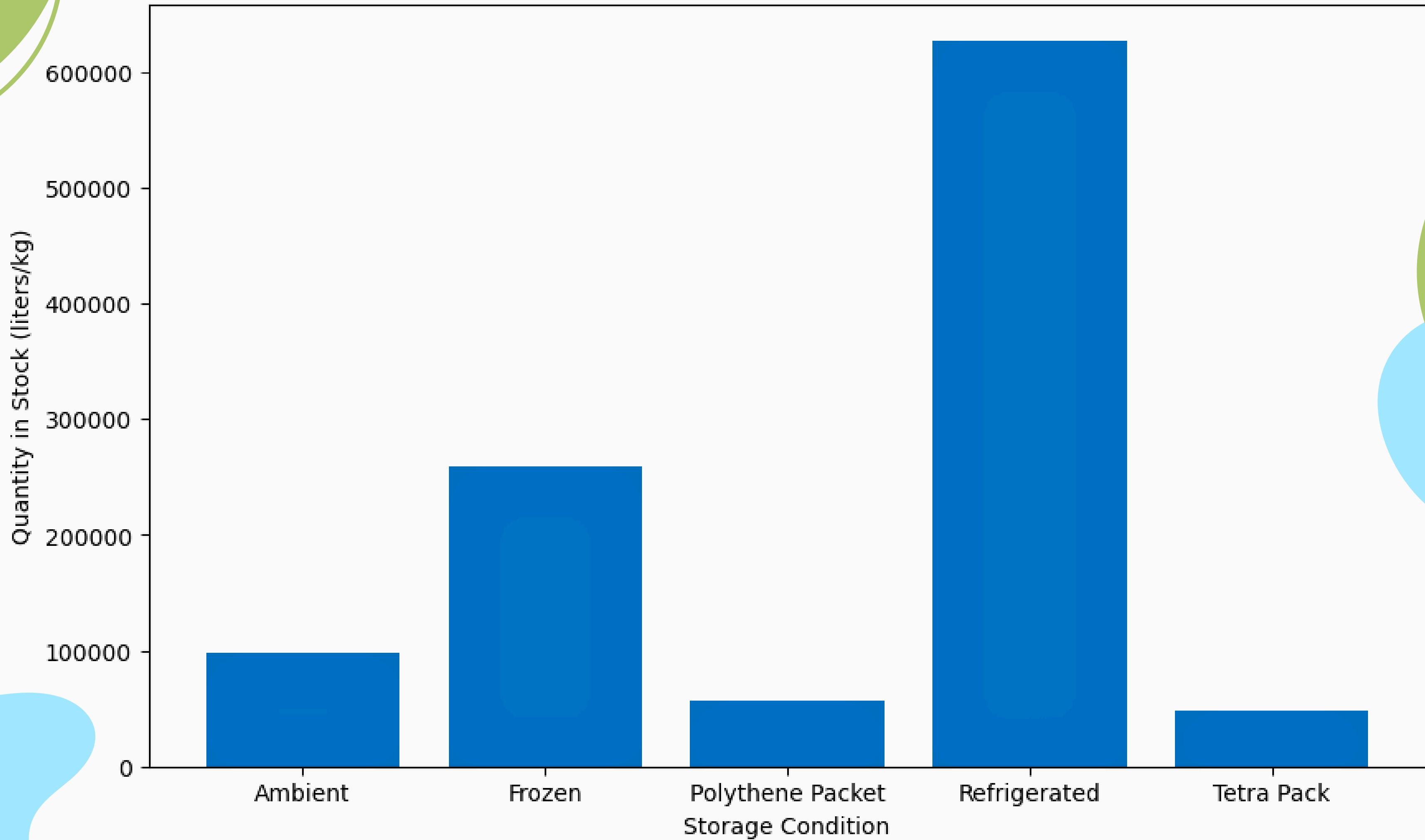




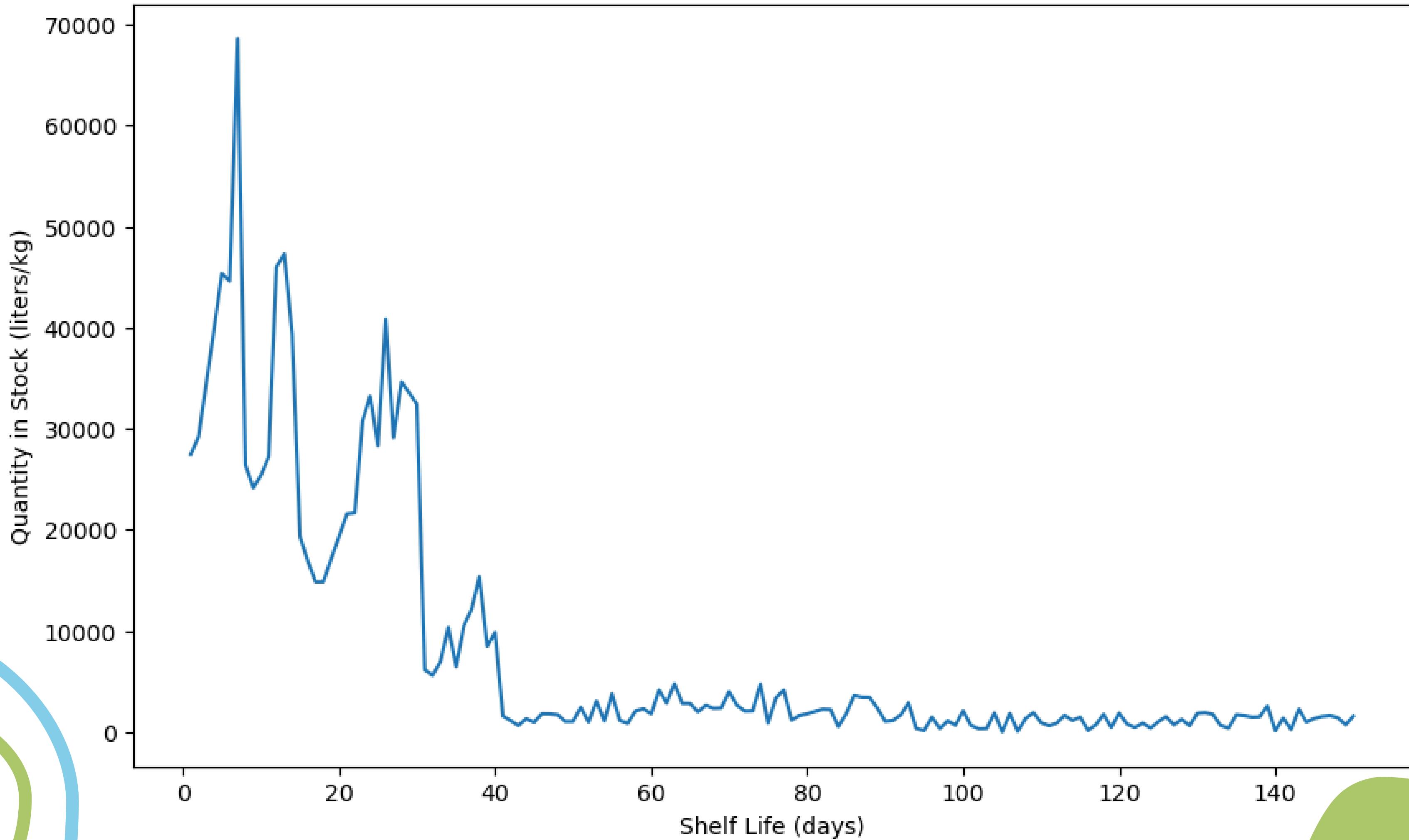
Total Revenue Trend for Each Brand Over the Years

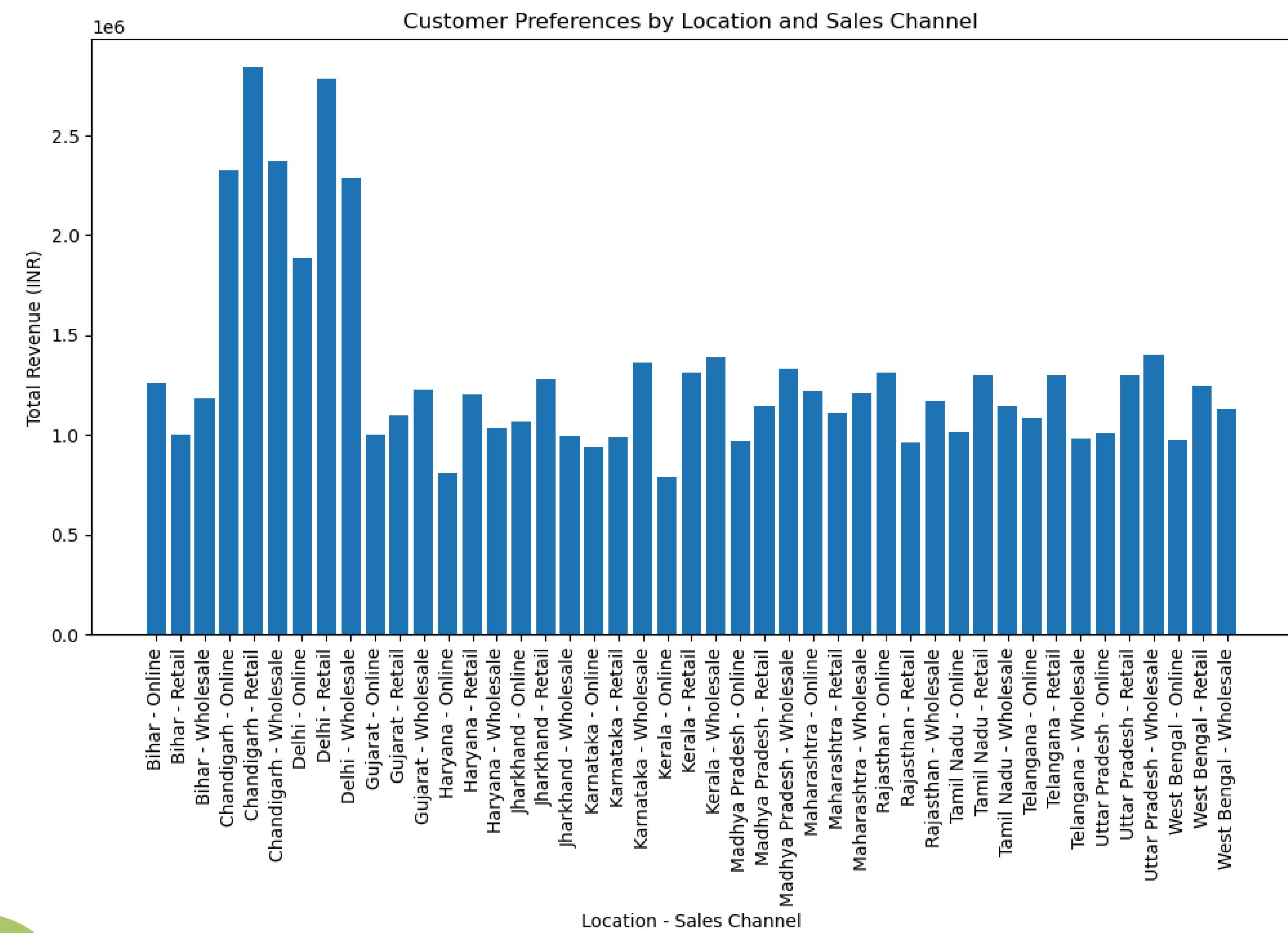


Impact of Storage Conditions on Quantity of Available Dairy Products

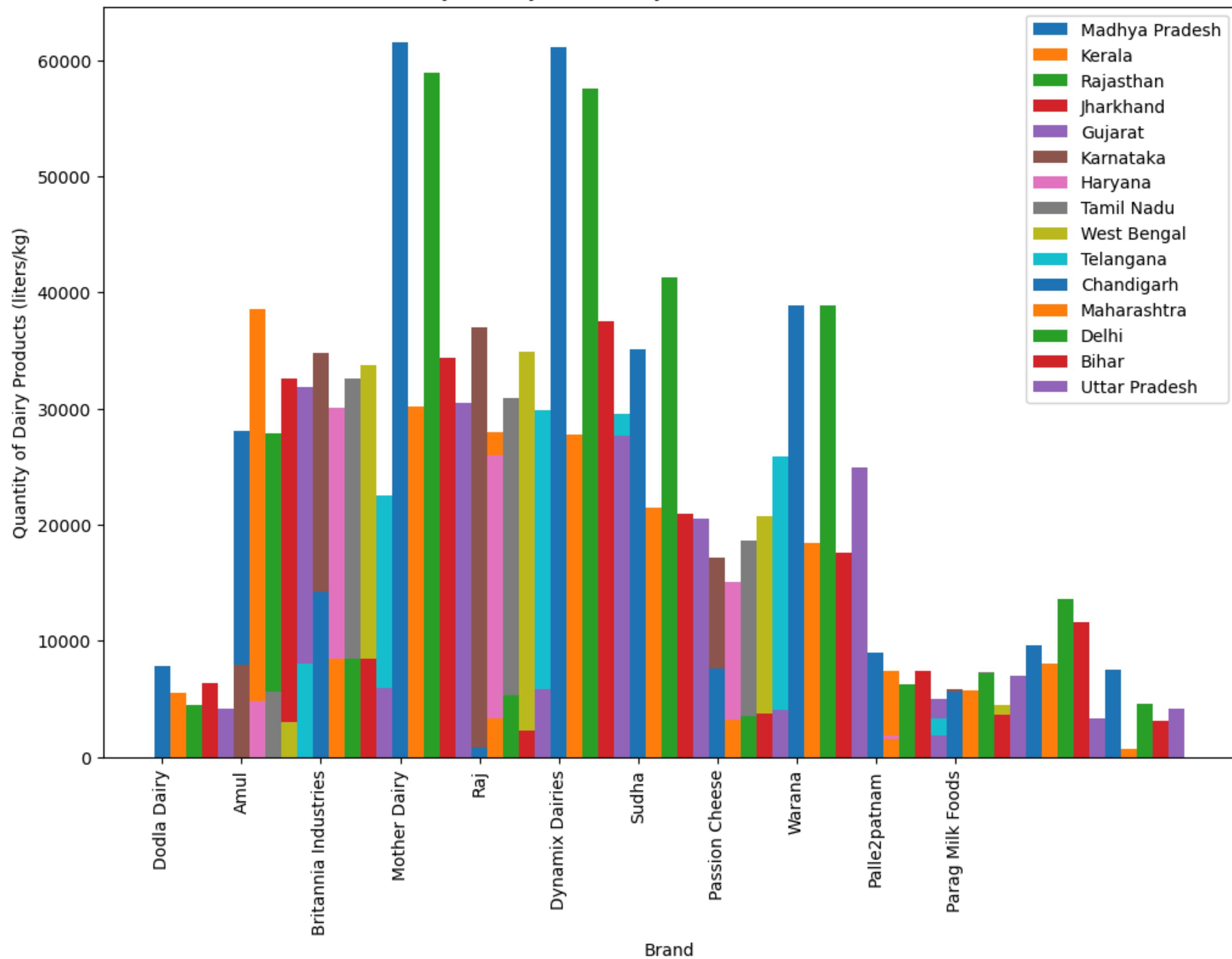


Impact of Shelf Life on Quantity of Available Dairy Products





Quantity of Dairy Products by Brand and Customer Location



Polynomial Regression Model Overview

What is Polynomial Regression?

Polynomial regression models the **relationship between the dependent and independent variables** as an **nth-degree polynomial**, allowing it to capture complex, non-linear relationships.

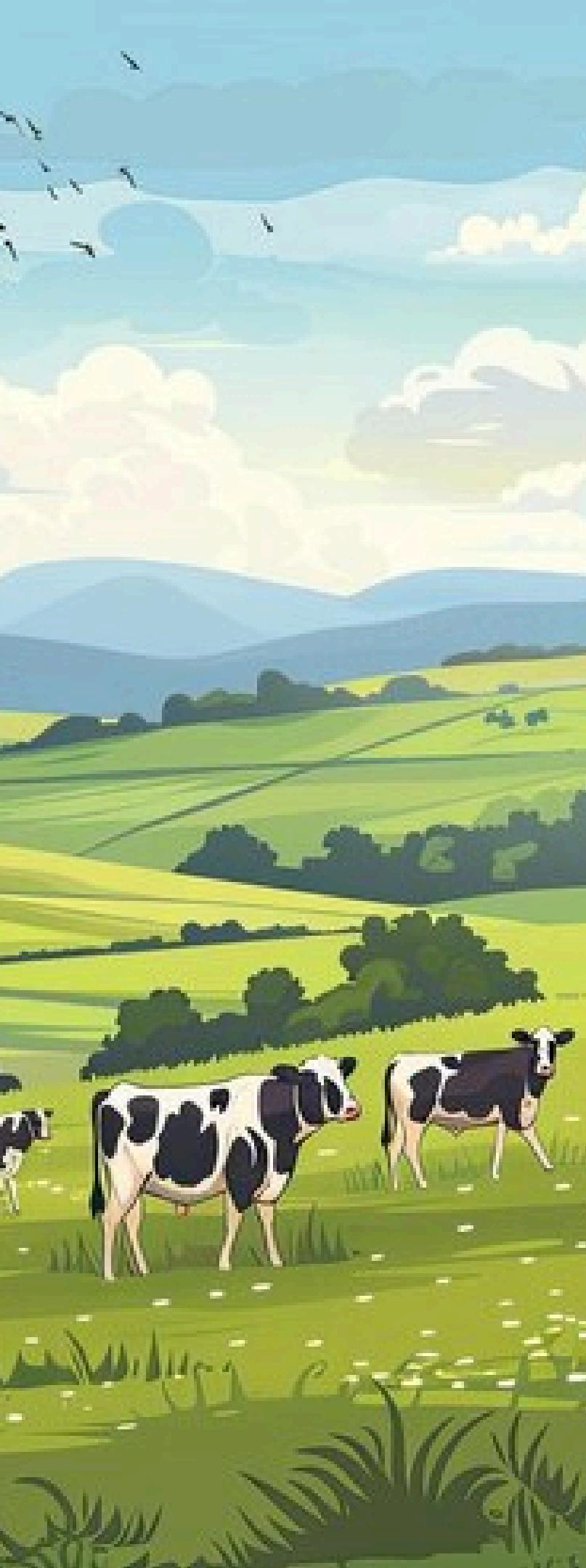
Application to Our Dataset

In our dairy dataset, polynomial regression predicts total revenue based on features like quantity sold, price per unit, and year. This method **effectively captures trends in revenue data that a linear model might miss**.

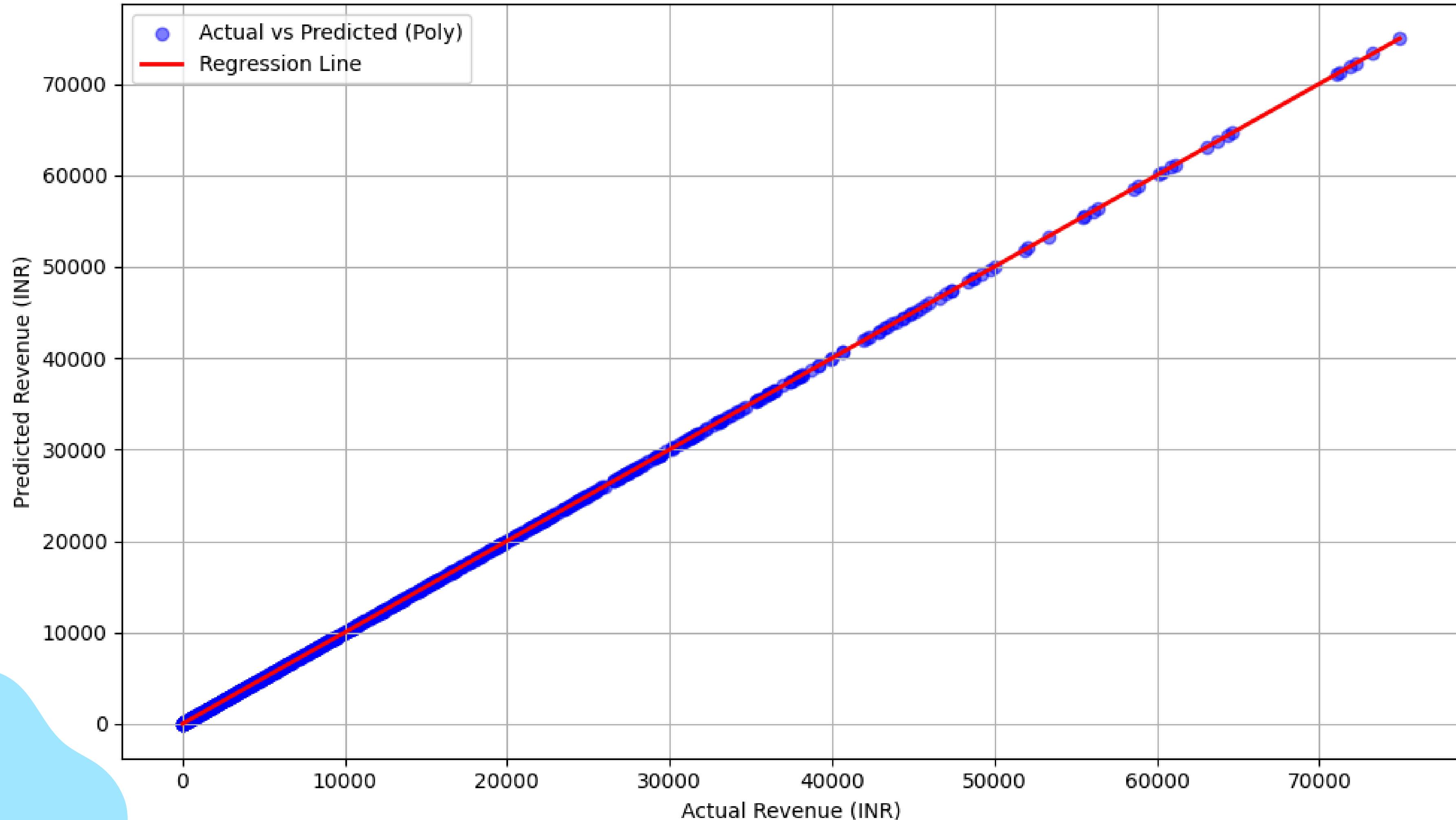
Model Performance

The **Mean Squared Error (MSE)** of our polynomial regression model is **7.368153174128262e-23**.

This **very low** value indicates that the **model's predictions are extremely close to the actual values**, demonstrating its accuracy and reliability in forecasting revenue trends in the dairy industry.



Actual vs Predicted Revenue with Polynomial Regression



T test

T-test Results:

Statistic: 0.5942818008508751

P-value: 0.5524795561906777

Null Hypothesis: There is no significant difference in the average price per unit between Milk and Yogurt.

Alternative Hypothesis: There is a significant difference in the average price per unit between Milk and Yogurt.

Conclusion: The t-test results indicate that the p-value is 0.5524795561906777

which is greater than the common significance level of 0.05. This suggests that we fail to reject the null hypothesis, meaning there is no significant difference in the average price per unit between Milk and Yogurt.



ANOVA

ANOVA Test Results:

Statistic: 1.6149214872627093

P-value: 0.1677820167340585

Null Hypothesis: There is no significant difference in the average price per unit among different dairy products.

Alternative Hypothesis: There is a significant difference in the average price per unit among different dairy products.

Conclusion:

The ANOVA test results show a p-value of 0.1677820167340585, which is greater than 0.05, indicating that we fail to reject the null hypothesis. This suggests there is no significant difference in the average price per unit among different dairy products.



Chi-square Test

Results:

Statistic: 5.664598626930544

P-value: 0.6847435405581676

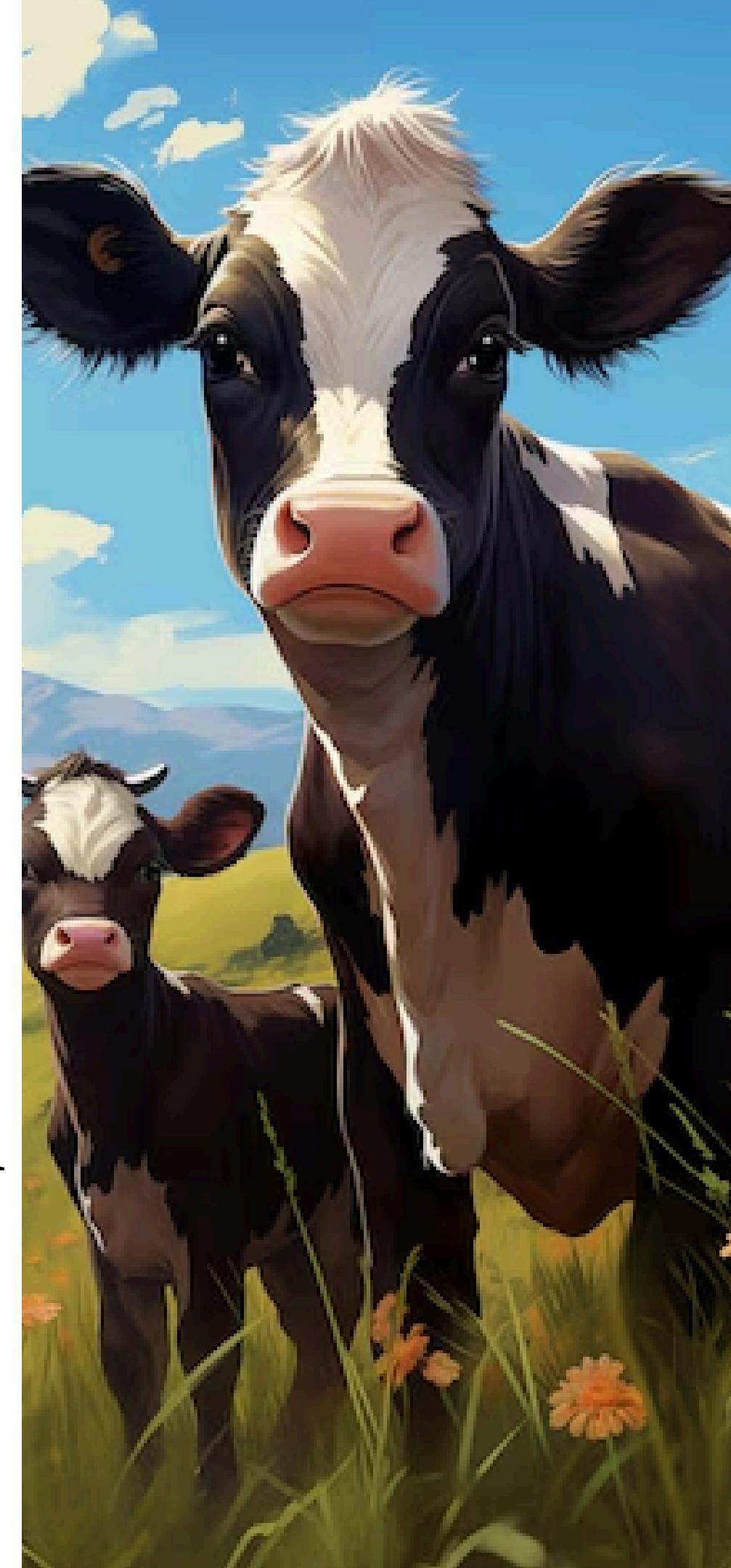
Degrees of Freedom: 8

Null Hypothesis: There is no association between Sales Channel and Storage Condition.

Alternative Hypothesis: There is an association between Sales Channel and Storage Condition.

Conclusion:

The Chi-square test results show a p-value of 0.6847435405581676 , which is greater than 0.05, indicating that we fail to reject the null hypothesis. This suggests there is no significant association between the Sales Channel and Storage Condition.



Z-test

Results:

Statistic: -22.016292887641388

P-value: 2.010537401822827e-107

Null Hypothesis: The proportion of products sold through Wholesale is equal to 50%.

Alternative Hypothesis: The proportion of products sold through Wholesale is not equal to 50%.

Conclusion:

The Z-test results show a p-value of 2.010537401822827e-107, which is significantly less than 0.05, indicating that we reject the null hypothesis. This suggests that the proportion of products sold through Wholesale is not equal to 50%.

The F-test is not applicable in this situation because it's typically used to compare variances between two populations, which isn't what we're trying to do with our dairy dataset.



Conclusion

1. The analysis conducted focused on price variations, sales channels, and revenue forecasting.
2. Statistical tests, including T-tests and ANOVA, showed no significant price differences among dairy products, while wholesale sales were confirmed to be less than 50%.
3. Polynomial regression effectively predicted revenue trends by capturing non-linear relationships between sales data and product attributes.
4. Overall, the analysis provides clear insights into pricing patterns, sales distribution, and future revenue forecasting for the dairy industry.



