

Business Report: Customer Segmentation and Sales Analysis

EDA | RFM | MBA - Sales_Data.xlsx



EDA Overview

- Used Python for initial analysis and visualization.
- Dataset: 3 years of transaction-level sales data.
- Objective: Identify trends, behaviors, and patterns in customer purchases.

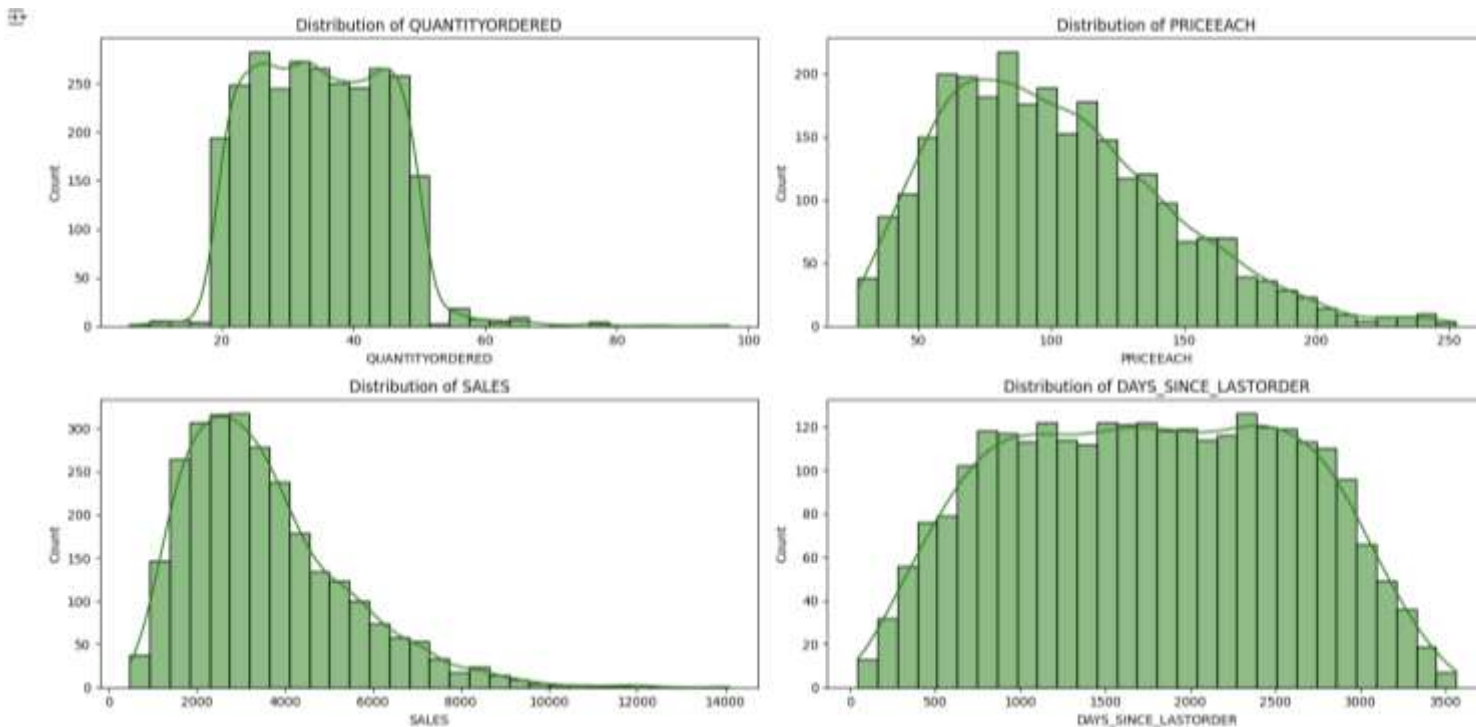
Summary Statistics

- Sales ranged from a few dollars to several thousand.
- Top selling product lines: Classic Cars and Motorcycles.
- USA, France, and Germany are top contributing countries.

	count	mean	min	25%	50%	75%	max	std
ORDERNUMBER	2747.0	10259.761558	10100.0	10181.0	10264.0	10334.5	10425.0	91.877521
QUANTITYORDERED	2747.0	35.103021	6.0	27.0	35.0	43.0	97.0	9.762135
PRICEEACH	2747.0	101.098951	26.88	68.745	95.55	127.1	252.87	42.042548
ORDERLINENUMBER	2747.0	6.491081	1.0	3.0	6.0	9.0	18.0	4.230544
SALES	2747.0	3553.047583	482.13	2204.35	3184.8	4503.095	14082.8	1838.953901
ORDERDATE	2747	2019-05-13 21:56:17.211503360	2018-01-06 00:00:00	2018-11-08 00:00:00	2019-06-24 00:00:00	2019-11-17 00:00:00	2020-05-31 00:00:00	NaN
DAYS_SINCE_LASTORDER	2747.0	1757.085912	42.0	1077.0	1761.0	2436.5	3562.0	819.280576
MSRP	2747.0	100.691664	33.0	68.0	99.0	124.0	214.0	40.114802

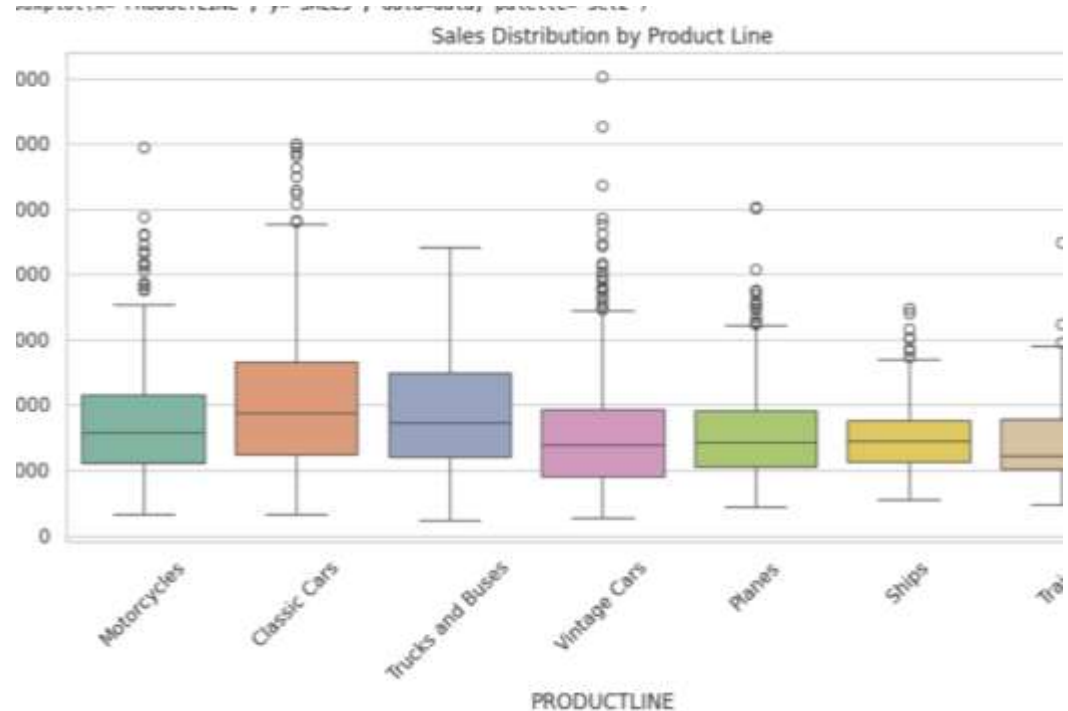
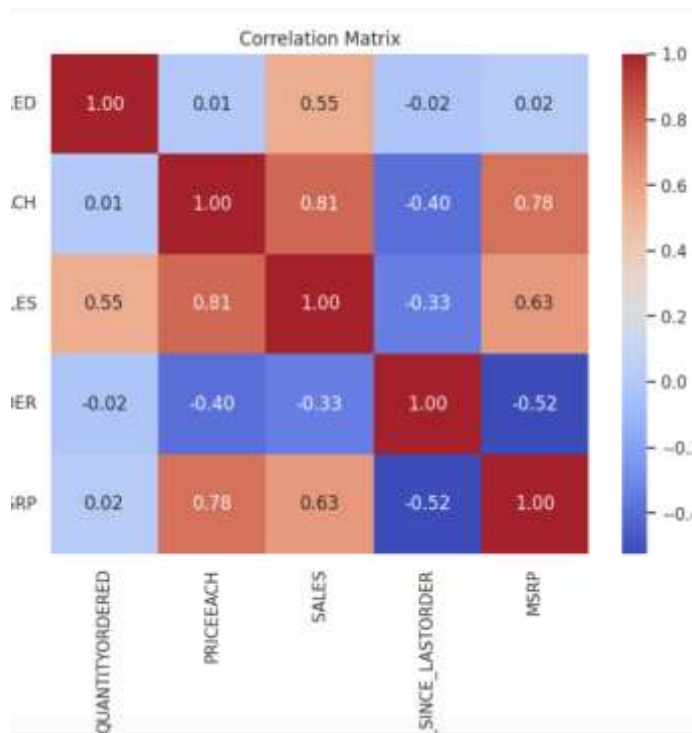
Univariate Analysis

- Sales show right-skewed distribution.
- Recency and frequency show expected drop-off patterns.
- Product Line has categorical variation in sales volume.



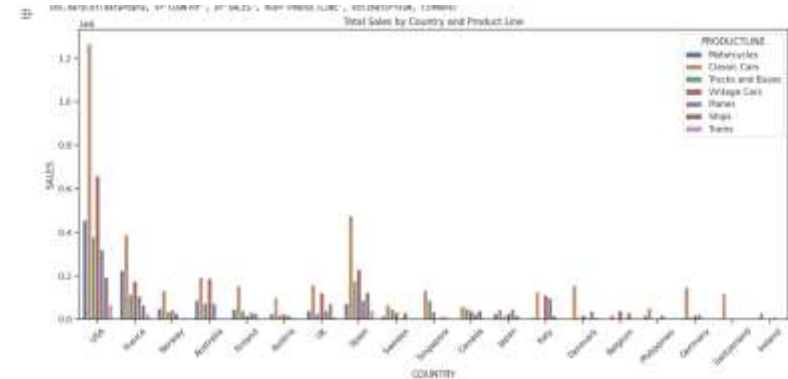
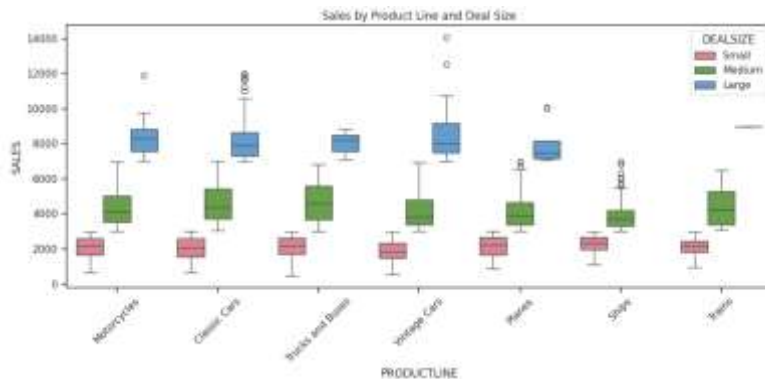
Bivariate Analysis

- Higher frequency correlates with higher monetary value.
- Classic Cars and Vintage Cars have high revenue per order.
- Seasonality evident in quarterly spikes.



Multivariate Analysis

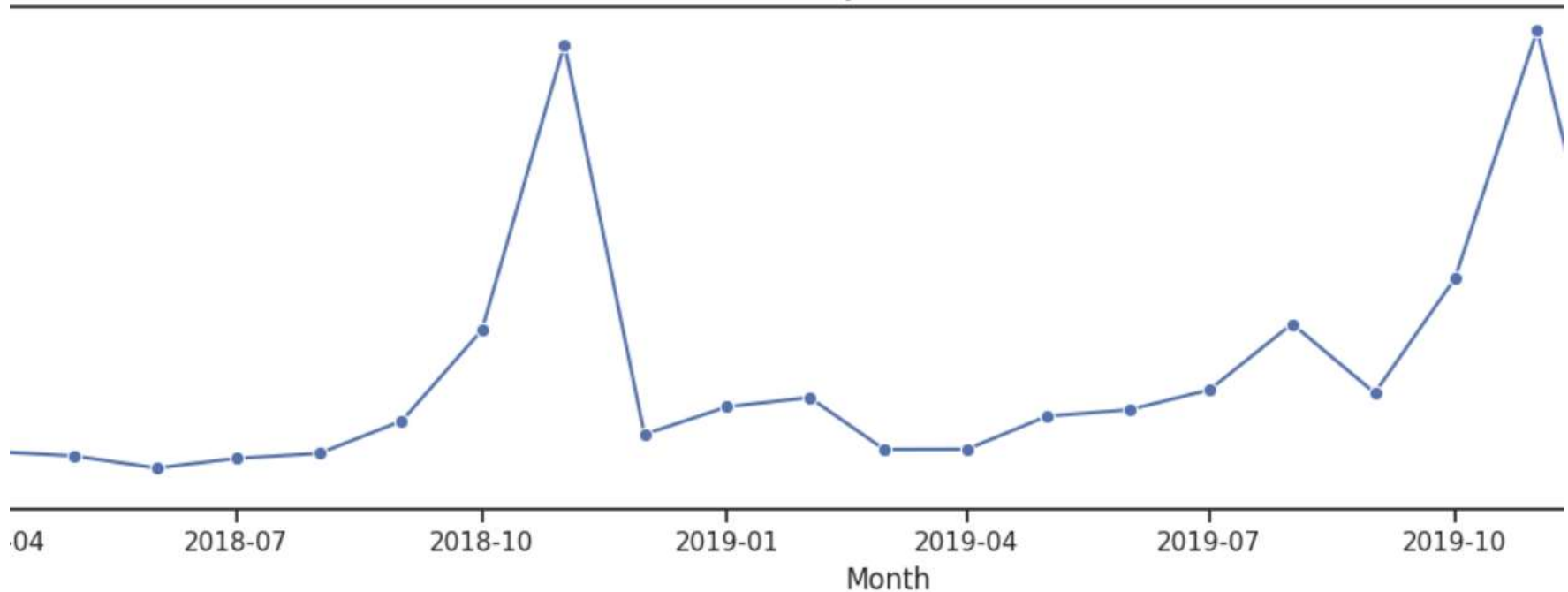
- Customers with high frequency also contribute more monetary value.
- Large deal sizes appear more in high-value orders.
- Outliers present for high Recency but low Frequency.



Time Series Trends

- Peak sales in Q4 (holiday season).
- 2020-2022: consistent quarterly trends with dips mid-year.
- Clear annual sales growth trend visible.

Monthly Sales Trend



Geographic and Category Insights

- Top cities: Madrid, Paris, San Francisco.
- Top categories: Classic Cars, Motorcycles.
- Certain categories only popular in specific regions.

Country vs Sales





Insights from EDA

- Focus on high-frequency, high-monetary customers.
- Target seasonal campaigns before Q4.
- Consider region-specific category promotions.

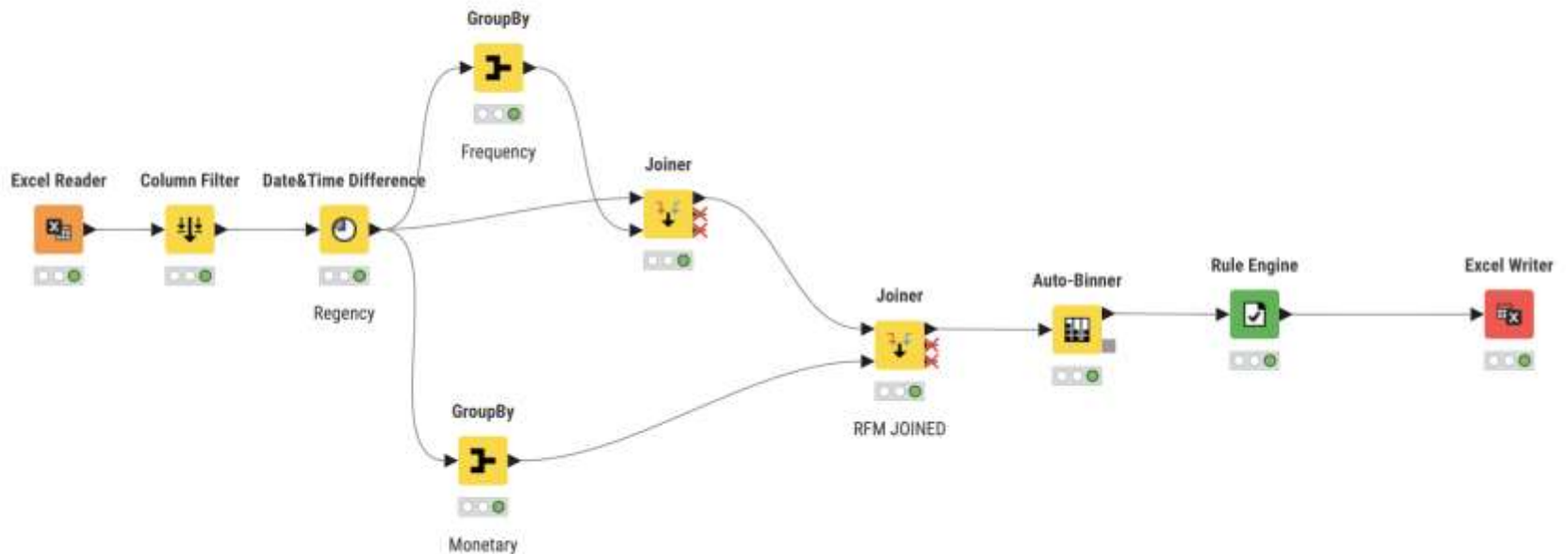


RFM Analysis Approach

- Used KNIME for Recency, Frequency, Monetary scoring.
- Auto-binner used to generate scores from 1 to 5 for each metric.
- Segmented customers using RFM scores.

RFM Workflow (KNIME)

- Step-by-step pipeline: GroupBy, Auto-Binner, Joiner, Rule Engine.
- Final table includes RFM scores and assigned segment.
- Workflow saved as .knwf file.



Customer Segments from RFM

- 5 key segments identified:
 - - Best Customers (R=5, F=5, M=5)
 - - Loyal Customers (High F, High M)
 - - Potential Loyalists (High F)
 - - At Risk (Low R)
 - - Lost Customers (R=1, F=1)



Best Customers (Top 5)

- High Recency, Frequency, and Monetary scores.
- Highly engaged, frequent, and valuable.
- Should receive VIP rewards and early offers.



Customers on Verge of Churning

- Low recency but good past monetary value.
- Need re-engagement campaigns and surveys.
- Important to act before full churn.

Loyal and Lost Customers

- Loyal: Stable frequency, moderate recency.
- Lost: No recent orders, very low frequency.
- Loyal - sustain through regular communication.
- Lost - cost-sensitive reactivation offers.