Final Project Data Analyst

Jihad Akbar

+62 813 3232 6785

https://www.linkedin.com/in/jihadakbr

jihadakbr@gmail.com

TABLE OF CONTENTS

01

Introduction

Jihad Akbar

"Transforming complexity into **clarity**, uncovering **insights**, and turning data into actionable strategies through **collaboration**."

Experience

Now–Oct 2024	**Machine Learning Annotator** PT Nomura Research Institute Indonesia
Sep–Feb 2024	**Data Annotator** CVAT.ai Corporation

Education

Now–Oct 2024	**Data Science Bootcamp** dibimbing.id
2020–2016	**Bachelor of Science in Physics** Universitas Sebelas Maret

02

Previous projects

What I Have Learned in the Bootcamp

New York City Taxi Trip Duration (Oct-Dec 2024)

- Predicted NYC taxi trip durations using 2016 data by defining business objectives, preprocessing data, engineering features like distance and speed, and addressing NYC-specific location constraints.
- Built scalable machine learning workflows using reusable transformer components with OOP principles in scikit-learn pipelines, achieving an RMSLE score of 0.55 with a LightGBM Regressor.

github.com/jihadakbr/new-york-city-taxi-trip-duration

Credit Card Fraud Detection (Sep-Oct 2024)

- Developed a fraud detection model using European credit card transaction data, addressing class imbalance and optimizing fraud detection accuracy by prioritizing the precision-recall AUC metric and adjusting the decision threshold.
- Fine-tuned models with Grid Search CV, achieving a PR AUC of 0.812 and F2 score of 0.794 using the Extra Trees Classifier, with a decision threshold set at 0.570 to balance accuracy and minimize false negatives.

github.com/jihadakbr/credit-card-fraud-detection

RFM Analysis for Customer Segmentation in Automobile Sales

03

Main Project

MAIN PROJECT OUTLINE
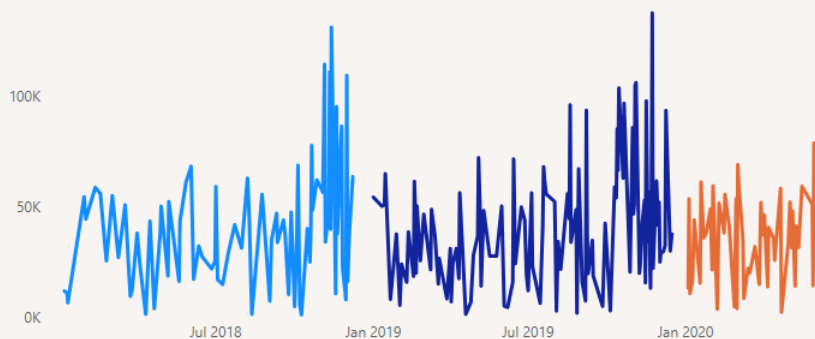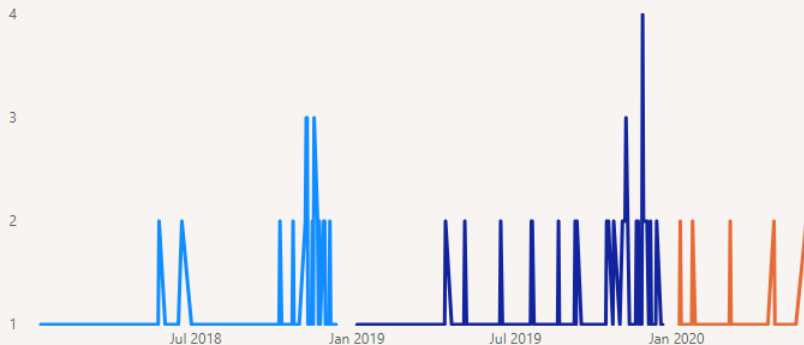
Project Background

Sales and customer engagement for the automobile company show fluctuating patterns, with significant peaks in November and December 2018 and 2019. These months are crucial for both sales and customer activity, but **consistency** remains a challenge.

Sales Trend

Total Customer Trend

Business Objective

Main Objective

This project aims to perform an RFM analysis on automobile sales data to segment customers by recency, frequency, and monetary values, identifying high-value clients. The insights and recommendations will enhance **marketing strategies**, improve **customer satisfaction**, and boost **sales**.

Specific Objective

To address the main objective, I have outlined specific objectives to achieve it:
1. What are the total numbers of customers, transactions, and sales, as well as the average values for recency, frequency, and monetary metrics?
2. Which customer segments contribute the most to total sales?
3. Which product categories generate the highest total sales across different customer segments, and what are the top-performing items within these categories for each segment?
4. How does the size of a deal correlate with total sales?
5. What are the trends in sales and the number of customers over time?

Data Understanding
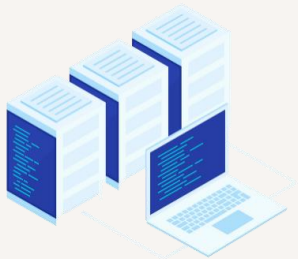
Collected from Kaggle, focusing on automobile sales.

Includes transaction details, customer demographics, and product information.
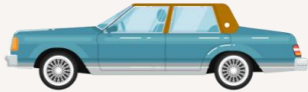
Covers transactions from 2018 to 2020.

Contains 2,747 transaction data and 20 columns.

Products

Motorcycles

Classic Cars

Vintage Cars

Trucks and Buses

Trains

Ships

Planes

Data Preprocessing

Step 1

Removing duplicates

⬇

No duplicate data found

Step 2

Handling missing values

⬇

No missing data found

Step 3

Converting data types

⬇

ORDERNUMBER (str) → int
ORDERDATE (str) → date

Step 4

Identifying and correcting errors

⬇

Formatted phone numbers and cleaned any non-alphabetical characters.

Data Preprocessing

Step 5
Creating a new "Profit" column

↓

Profit =
Sales – (Quantity Ordered
* MSRP)

Step 6
Detecting and
Handling Outliers

↓

No extreme
outliers found

Step 7
Transforming data for
RFM Analysis

↓

Created 9 customer
segments

Step 8
Saving preprocessed
data for the dashboard

↓

CSV of cleaned data
and
CSV of cleaned
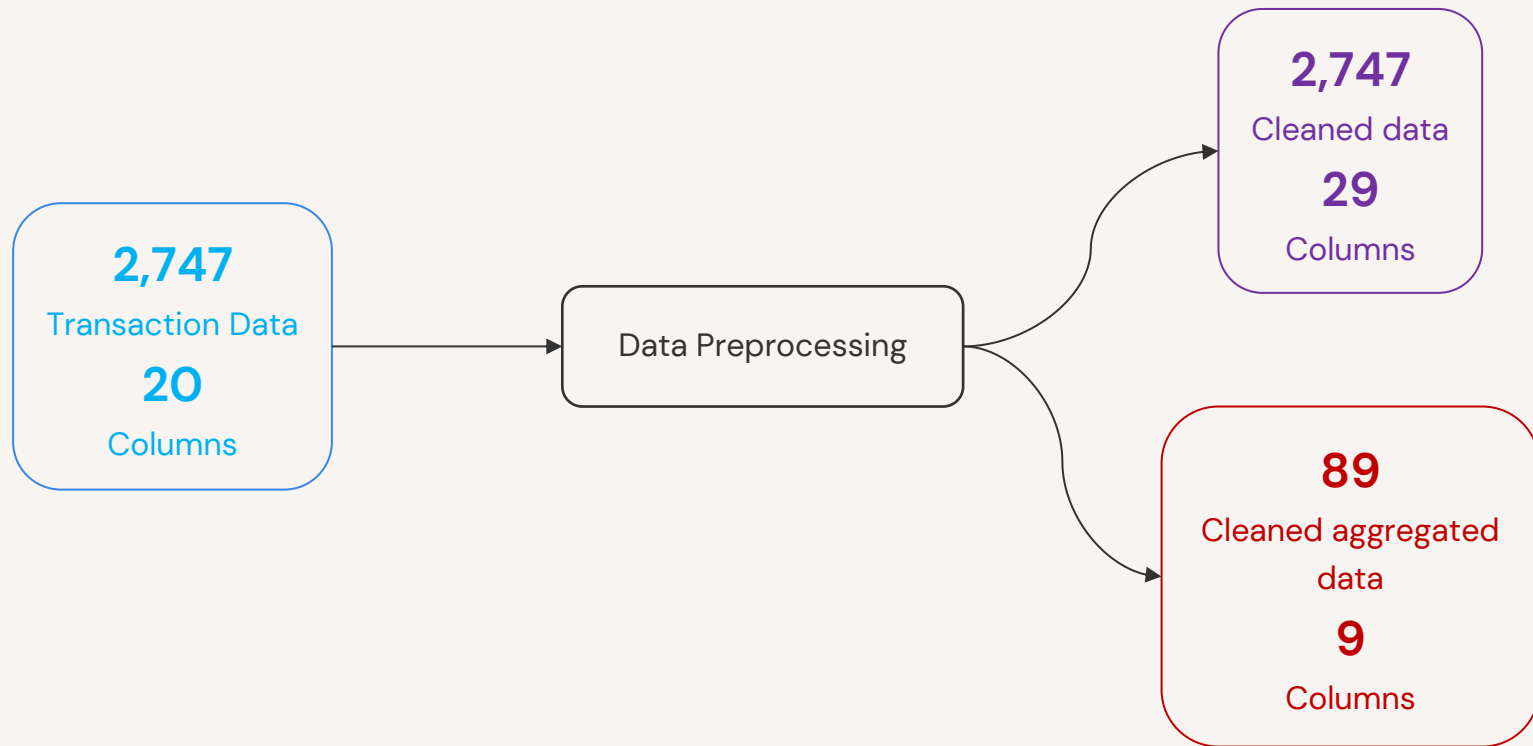aggregated data

Data Preprocessing Result

RFM Definition

Recency

How recently a customer made a purchase. The more recent the purchase, the more likely they are to buy again.

Frequency

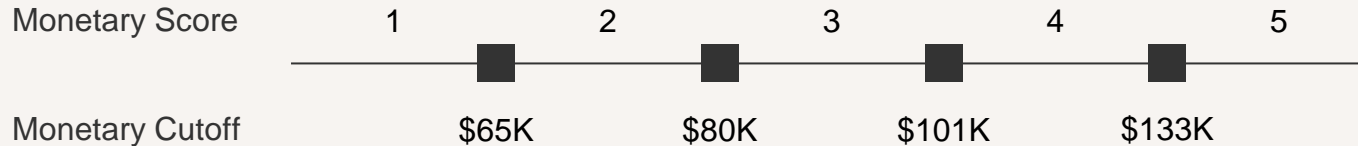How often a customer makes a purchase. The more often they buy, the more engaged they are with our brand.

Monetary

How much money a customer spends. The more they spend, the more valuable they are to our business.

Findings and Results

Recency Score	5	4	3	2	1

Recency Cutoff	55 days	146 days	199 days	259 days

Frequency Score	1	2	3	4	5

Frequency Cutoff	2x	3x	3x	4x

Monetary Score	1	2	3	4	5

Monetary Cutoff	$65K	$80K	$101K	$133K

Findings and Results

No	Segment	R	F	M	Description	Recommendation
1	Champion	4-5	4-5	4-5	Top customers who make frequent purchases, spend the most, and shop recently.	Offer exclusive discounts or VIP programs to maintain engagement.
2	Emerging Champions	3-5	3-5	3-4	Customers showing potential to become Champions but with slightly lower spending.	Provide personalized incentives (e.g. targeted discounts) to increase spending.
3	Enthusiastic Spenders	4-5	1-3	4-5	Recent shoppers who spend significantly but purchase less frequently.	Implement loyalty programs or subscription services to foster repeat buying.
4	Loyal Costomers	2-4	3-5	1-5	Customers who shop frequently with varying spending amounts.	Offer rewards and tiered loyalty benefits for continued engagement.
5	High-Potential Customers	2-4	1-2	3-5	High spenders with lower frequency of purchases, but potential for growth.	Provide targeted content and incentives to encourage more frequent purchases and deeper engagement.
6	Value Seekers	3-5	2-3	1-3	Regular shoppers who prefer budget-friendly options.	Highlight budget-friendly products, discounts, and time-limited promotions.
7	Casual Shoppers	3-5	1-2	1-3	Infrequent shoppers with minimal spend but recent engagement.	Use targeted offers to bring them back, like "special offer just for you!"
8	Premium Shoppers	5	4-5	1-2	Recent and frequent shoppers, but spend less than top-tier segments.	Encourage higher-value purchases with personalized offers or loyalty bonuses.
9	At Risk	1-2	1-5	1-5	Inactive customers who need re-engagement strategies.	Send "we miss you" messages with time-sensitive offers to bring them back.

Findings and Results

1. What are the total numbers of customers, transactions, and sales, as well as the average values for recency, frequency, and monetary metrics?

Total Customers	Total Transactions	Total Sales
89	298	$9.76M

Recency Average	Frequency Average	Monetary Average
180.61	3.35	$109.67K

Insight

- **Moderate Recency (Score 3)**: The 6-month (180-day) gap suggests a need for reactivation efforts.
- **Moderate Frequency (Score 3 or 4)**: Customers purchase 3–4 times on average, indicating room to boost engagement.
- **High Monetary Value (Score 4)**: Each customer contributes significantly to total sales; retaining them is critical.

Findings and Results

**2. Which customer segments contribute the most to total sales?**

Sales by Segment

Segment	Sales
Champions	$3.7M
At Risk	$1.7M
Loyal Customers	$1.7M
Emerging Champions	$1.3M
High-Potential Customers	$0.6M
Value Seekers	$0.3M
Premium Shoppers	$0.1M
Enthusiastic Spenders	$0.1M
Casual Shoppers	$0.0M

Low Impact

Total Customers by Segment

Segment	Total Customers
At Risk	25
Champions	18
Loyal Customers	16
Emerging Champions	14
High-Potential Customers	6
Value Seekers	6
Premium Shoppers	2
Casual Shoppers	1
Enthusiastic Spenders	1

Low Impact

Insight

- **Top Contributors**:
  - **Champions**: $3.7M (38.4% of sales, 18 customers) – High–value, frequent, and recent buyers.
  - **At Risk**: $1.74M (17.8%, 25 customers) – Significant contribution but need reactivation.
  - **Loyal Customers**: $1.69M (17.3%, 16 customers) – Consistent contributors.
- **Low Impact**:
  - **High–Potential Customers**, **Value Seekers**, **Premium Shoppers**, **Casual Shoppers**, and **Enthusiastic Spenders** contribute minimally to sales and customer count.

Findings and Results

3. Which product categories generate the highest total sales across different customer segments, and what are the top-performing items within these categories for each segment?


Sales by Product and Segment
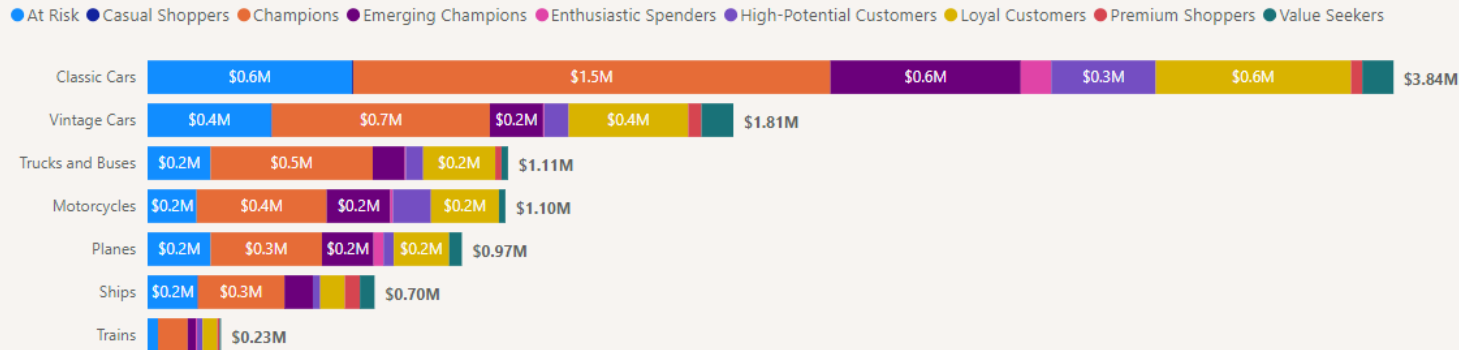
Insight

- **Top Categories by Sales:**
  - **Classic Cars** ($3.84M) lead in total sales, followed by **Vintage Cars** ($1.81M).
  - **Trucks and Buses** ($1.11M) and **Motorcycles** ($1.10M) also perform well.
- **Top-Performing Items by Segment:**
  - **Champions** lead in all product categories.
  - **At Risk** customers are in second place across all products.

Findings and Results

**4. How does the size of a deal correlate with total sales?**

Relationship Between Deal Size and Average Sales
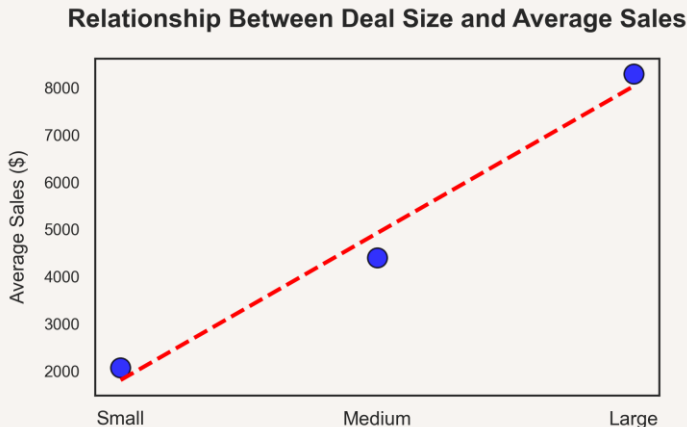
Insight
- **Positive Correlation:** Larger deals generate significantly higher average sales (Large: $8,283, Medium: $4,397, Small: $2,063).
- **Stepwise Growth:** Average sales nearly double from Small to Medium and Medium to Large, showing consistent growth with deal size.

Findings and Results

5. What are the trends in sales and the number of customers over time?

Sales Trend

Total Customer Trend
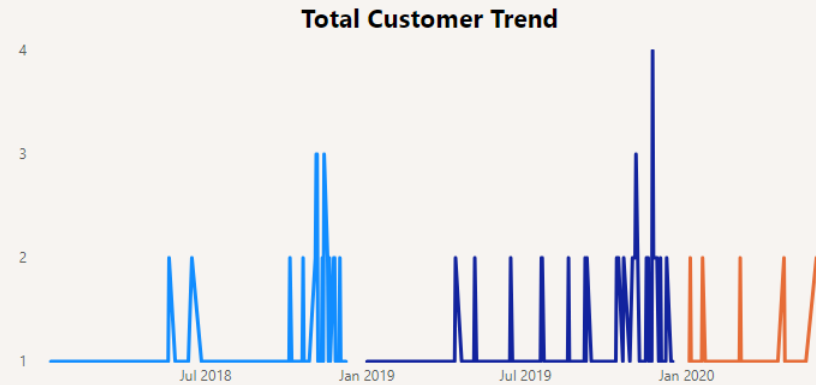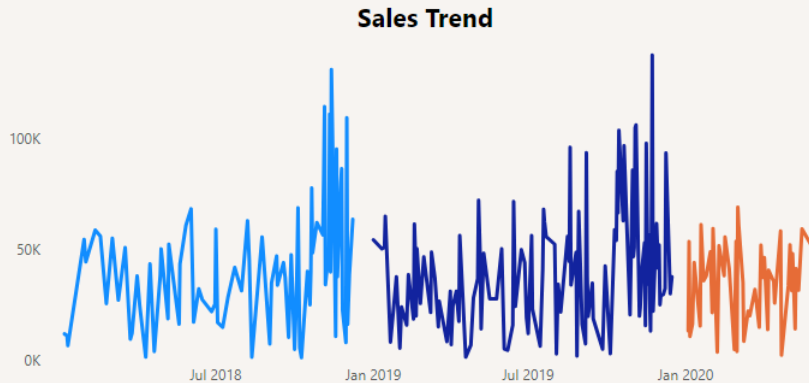
Insight

- **Sales Trends**: Sales show fluctuating patterns with significant peaks in November and December 2018 and 2019, indicating strong seasonal activity.
- **Customer Trends**: Customer counts are highly variable, with spikes in November and December 2018 and 2019, suggesting engagement linked to specific events or campaigns.
- **Key Periods**: November and December 2018 and 2019 are crucial for both sales and customer engagement, highlighting the importance of seasonal marketing strategies.

Recommendations

Based on my findings, here are my recommendations to drive sales growth and enhance customer engagement:

1. **Boost Customer Retention for Moderate Recency & Frequency Segments**
   - Target customers with a 6-month gap and 3-4 purchases for reactivation. Offer personalized promotions and loyalty programs to reduce recency and increase purchase frequency.
2. **Prioritize High-Value Customer Segments**
   - Focus marketing efforts on "Champions" and "At Risk" segments, which contribute the most to **total sales**. Implement VIP programs, exclusive discounts, and tailored incentives to strengthen relationships.
3. **Focus on "Champions" and Reactivating "At Risk" Customers**
   - Retain and reward Champions, and target campaigns to reactivate At Risk customers, as both are the highest contributors across **all product categories**. Additionally, focus on Classic Cars and Vintage Cars, which are the top sales categories, driving the largest portions of revenue in these segments.
4. **Encourage Larger Deals for Higher Sales**
   - Increase sales by offering incentives for bulk purchases, volume discounts, and additional perks. Larger deals contribute significantly more to revenue, so capitalize on this trend.
5. **Capitalize on Seasonal Peaks**
   - Based on the analysis of November and December 2018 and 2019, focus on these months with holiday promotions, limited-time offers, and events to maximize sales and engagement.
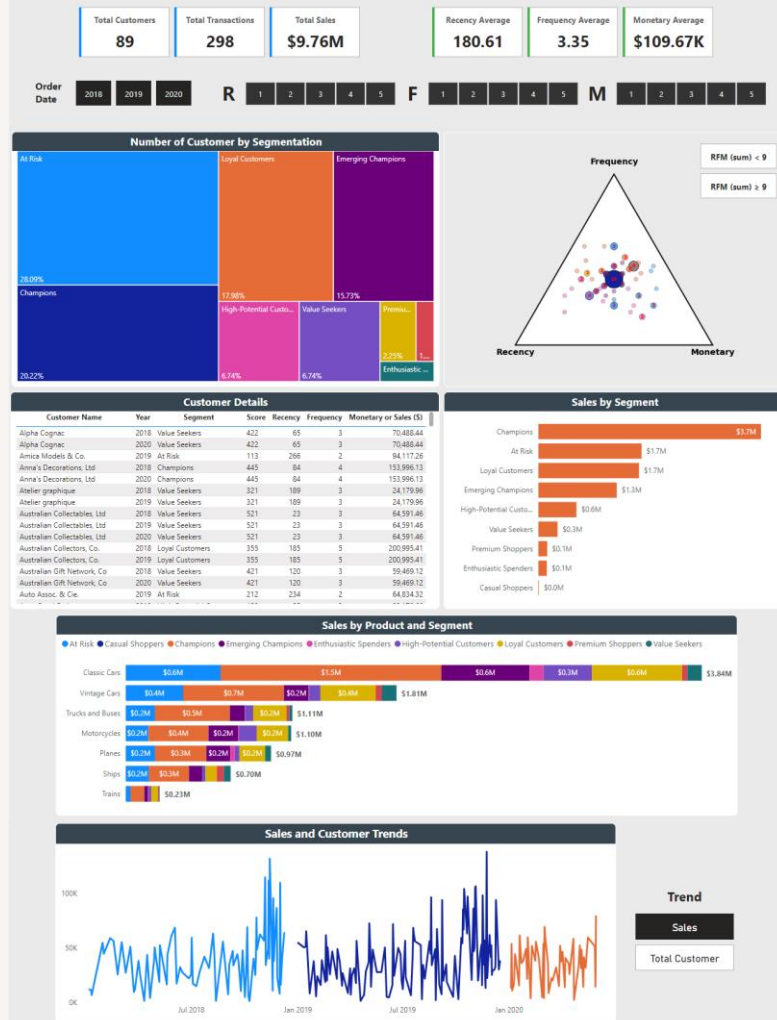
Customer Dashboard

To maintain and monitor the customer segment, I have provided dashboard for the related stakeholder to access

Link to the Dashboard:

[G-Drive](#)

04

Appendix

Link

Dashboard: RFM_DS29_Dashboard_Jihad_Akbar.pdf

Preprocessing: Final_DA_DS29_Jihad_Akbar.ipynb

THANK YOU

Do you have any questions?

Jihad Akbar

+62 813 3232 6785

https://www.linkedin.com/in/jihadakbr

jihadakbr@gmail.com