Car Shipment Delay Prediction

Predicting delays and financial losses in car part shipments using Machine Learning

AMNEET KAUR | Data Scientist |

Problem Statement

Business Problem

- Car manufacturers often face unpredictable shipment delays due to various reasons like supplier issues, weather, holidays, and strikes. These delays can lead to production slowdowns and significant financial losses.

Project Goal

To build a system that can:

- **Predict** if a car shipment will be delayed
- **Estimate** how many days it might be delayed
- **Calculate** the potential **financial impact** of that delay

Why This Matters

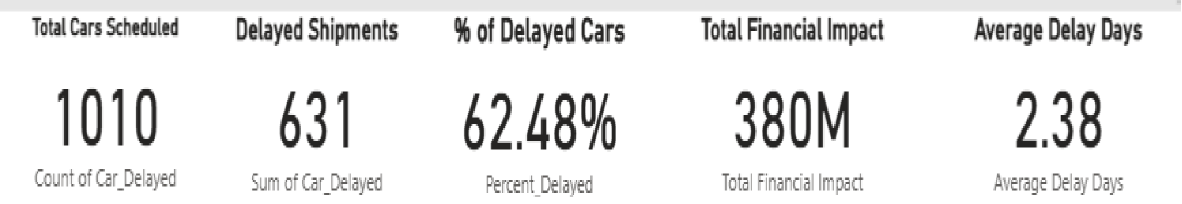
By identifying delays early, the company can take **proactive steps** like adjusting schedules or sourcing from alternate suppliers to reduce losses, plan better, and improve overall supply chain efficiency.

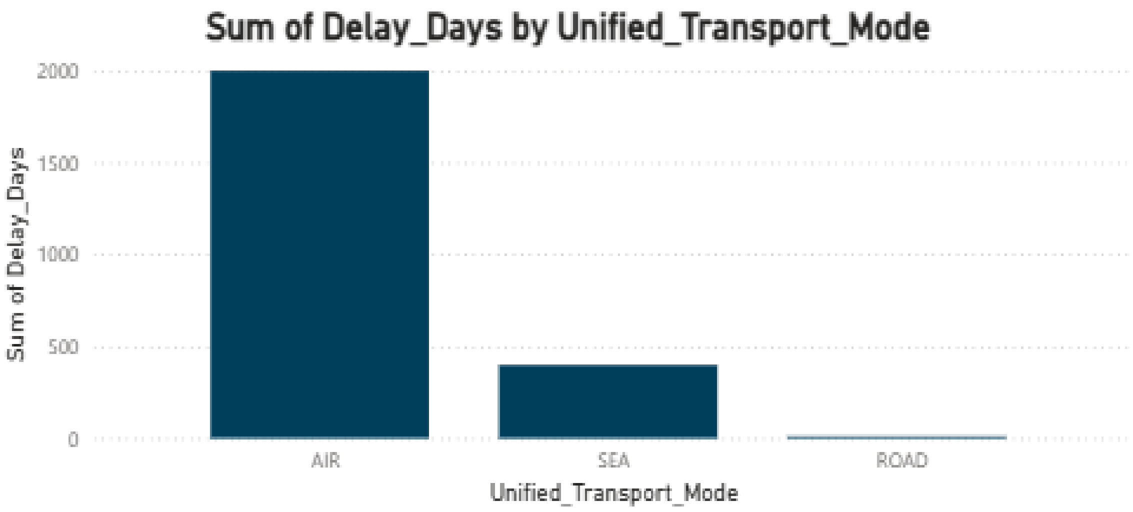
Understanding the Dataset

- I used an AI-generated dataset with ~1,000 car shipments, including part delivery dates, supplier info, delay reasons, and external factors like weather, strikes, and holidays.
- To avoid data leakage, I excluded Supplier_Reliability, which could unintentionally leak future information. Instead, I used historical Supplier_Performance_Score.
- After cleaning and feature engineering, I used:
- 15 features for delay prediction and delay days models
- 5 features for financial impact prediction

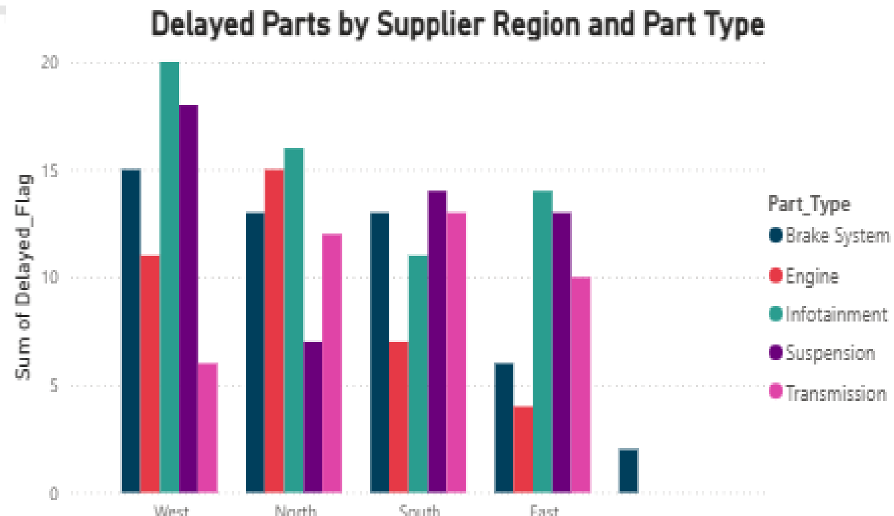
her_(	Strike_Not	Productio	Valid_Dela	Total_Part	Route_Typ	Distance_	Traffic_Le	Supplier_F	Contract_I	Engine_Su	Transmiss	Brake_Sys	Suspensic	Infotainm	Engine_Re	Transmiss	Brake_Sys	Suspensic	Infotainm	Traffic_Se
n	0	Morning	TRUE	1	Rural	227.2142	Low	0.12	250	East_Roac	North_Roa	South_Sea	East_Sea	North_Roa	0.302632	0.329114	0.282609	0.23913	0.310345	1
	0	Morning	FALSE	1	Urban	29.08404	Low	0.98	250	North_Sea	East_Air	East_Sea	North_Air	West_Sea	0.289474	0.328571	0.230769	0.311688	0.4	1
	0	Morning	FALSE	0	Mixed	99.05497	Low	0.75	100	West_Air	East_Air	East_Air	West_Roa	East_Air	0.346154	0.328571	0.383562	0.356322	0.298851	1

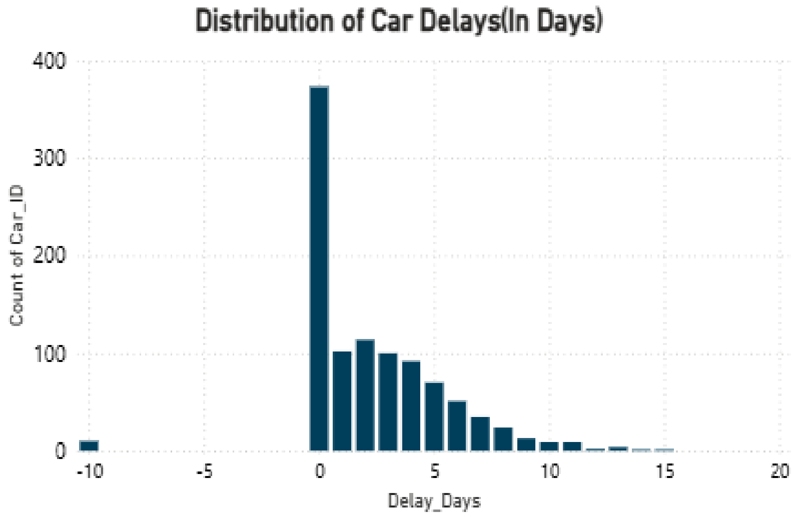
Exploratory Insights into Delay Patterns

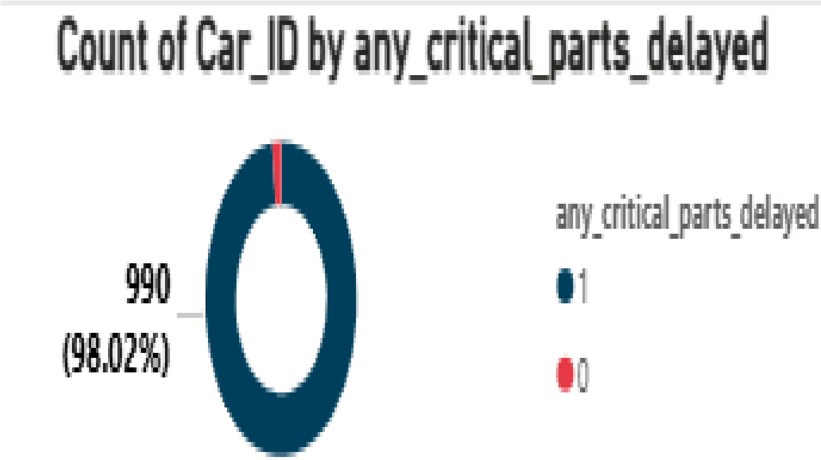




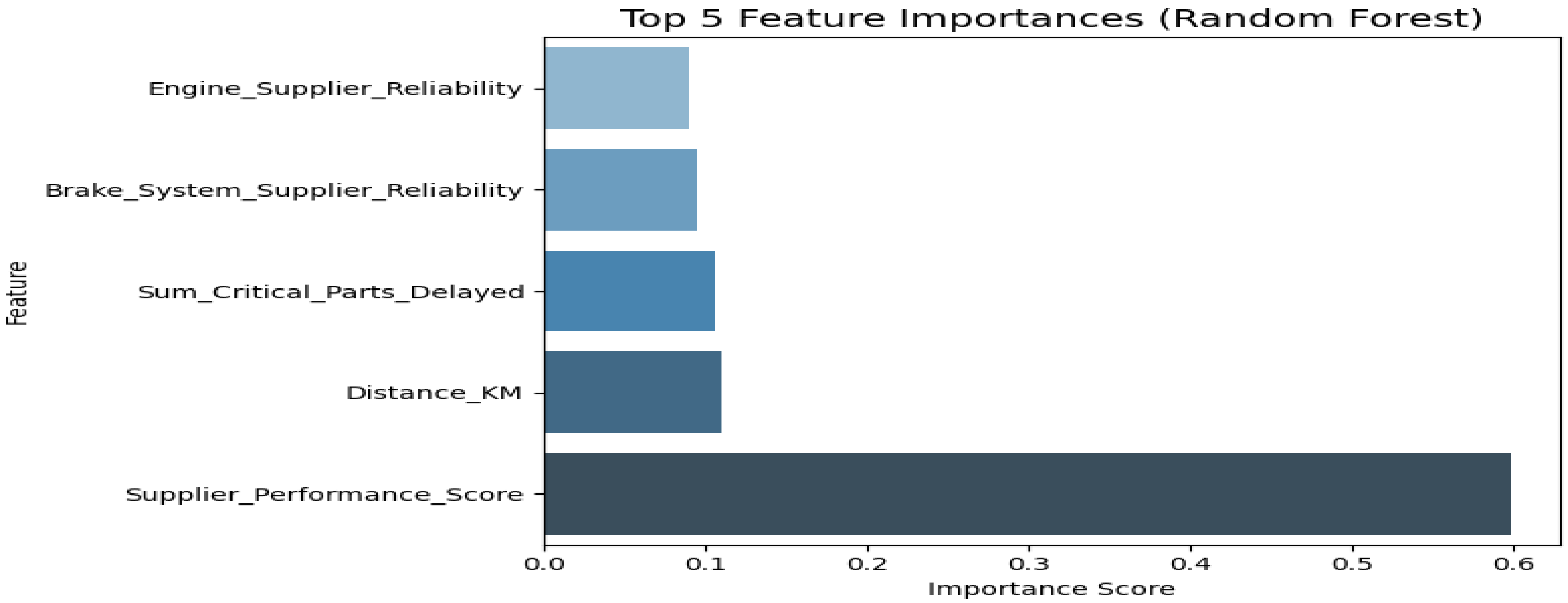
- Unreliable transport methods like **air** showed significantly higher cumulative delay days.
- Certain regions and part types — **especially Infotainment and Suspension parts** from **west zones** — contributed heavily to delays.
- **Most delays** occurred when **critical components** were **delayed**.
- Majority of delayed cars were late by **0-5 days**, but there's a long tail with extreme delays of **5-15 days**.
- **62.48%** of total cars experienced shipment delays, causing **380 Million** of total financial impact.





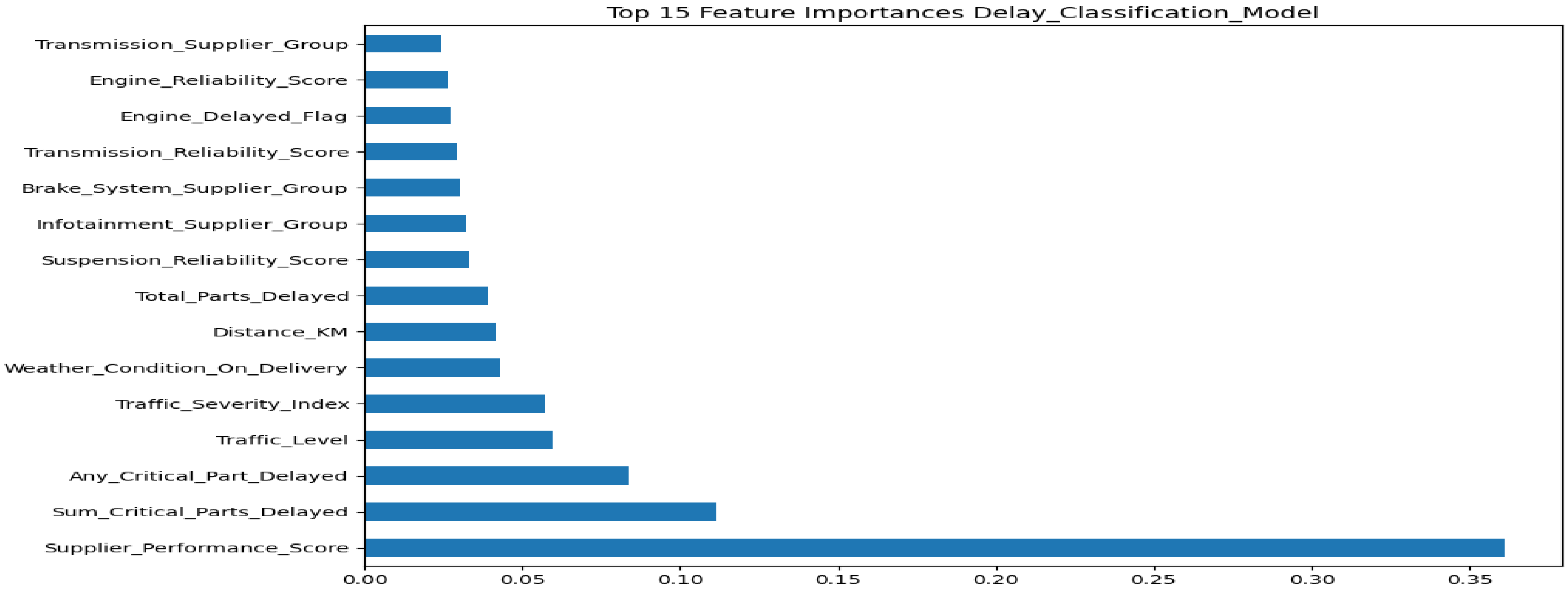


Feature Importances of Random_Forest_Financial_Impact



Supplier_Performance_Score and Distance_KM drive cost the most.

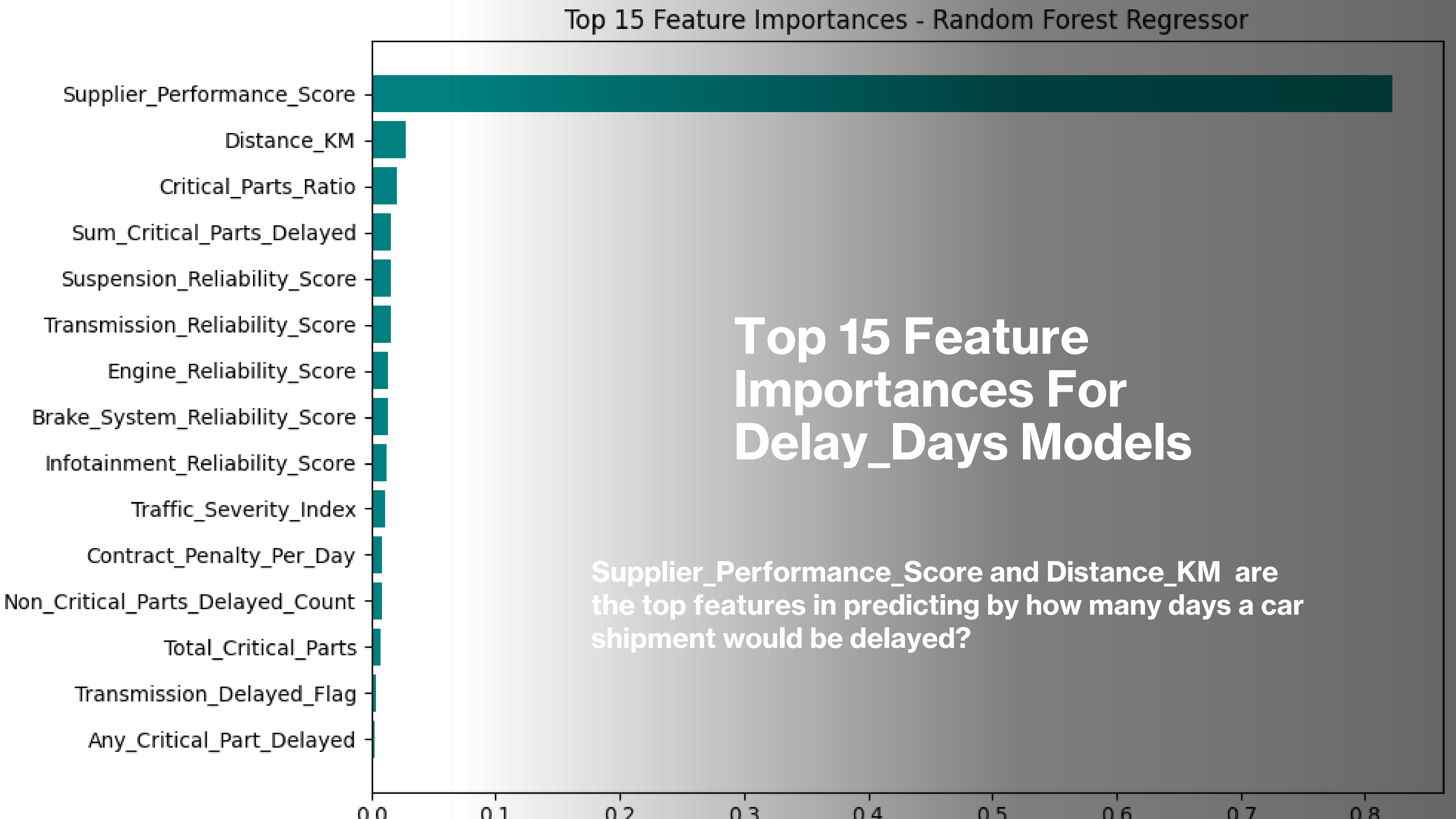
Top 15 Features Used in Classification Model



Supplier_Performance_Score and Sum_Critical_Parts_Delayed are the top features in classifying whether a car shipment would be delayed yes or no, with probability.

Top 15 Feature Importances For Delay_Days Models

Supplier_Performance_Score and Distance_KM are the top features in predicting by how many days a car shipment would be delayed?



Modeling Approach

Classification Model

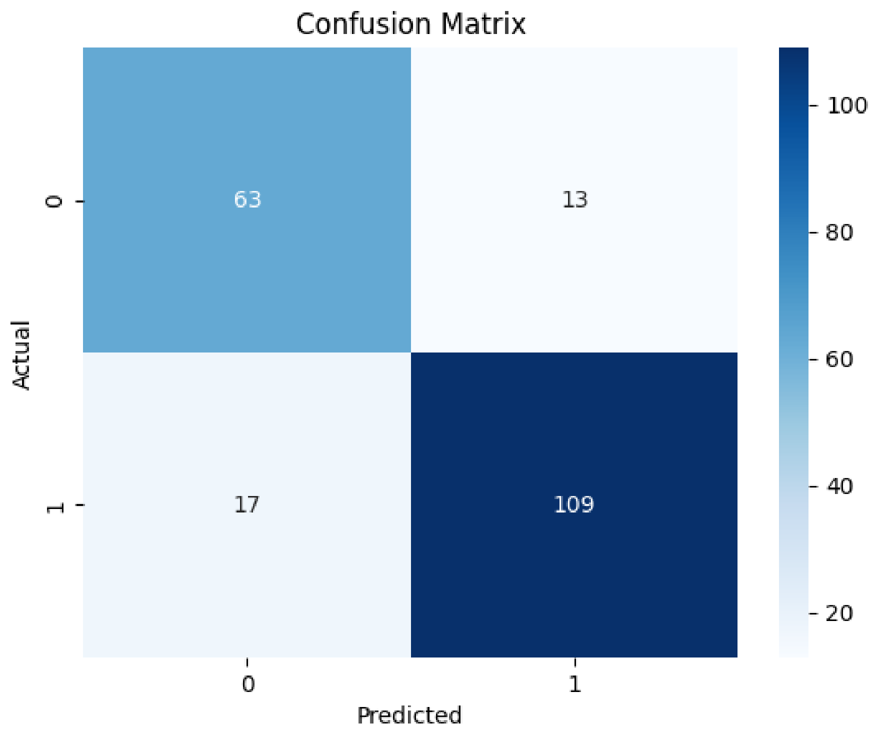
Regression - Delay Days

Regression - Financial Impact

I built **three machine learning models** to tackle different business questions:

- **Will the shipment be delayed?** (Classification – Random Forest)
  If delayed, by how many days? (Regression – Random Forest)
  What's the financial impact? (Regression – Random Forest)
- For each model, I selected the top-performing features after preprocessing, label encoding, and handling missing values.
- Models were evaluated using metrics like **F1-score**, **MAE**, and **R^2**, depending on the task.

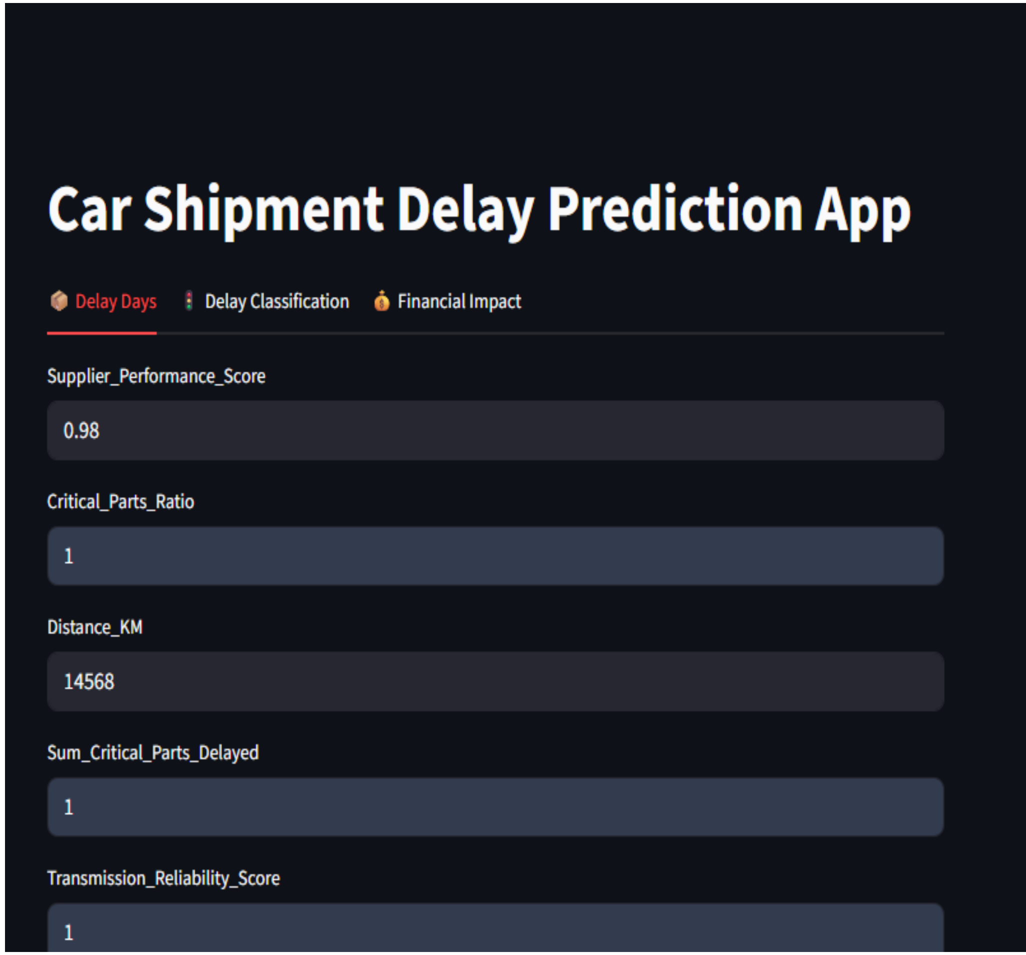
Model Performance Summary



Model	Key Metric(s)	Performance
Delay Classification ✅	Accuracy / F1	85% / 0.88
Delay Days (Regressor) 📅	MAE / R^2	0.86 / 0.81
Financial Impact 💵	MAE / R^2	$445 / 0.73

Interactive Delay Prediction App

- I built a Streamlit web app that allows business users to:
- 🔍 **Predict** if a car shipment will be delayed (Yes/No)
- 📅 **Estimate** the number of delay days
- 💸 **Quantify** the financial loss due to the delay
- The app uses 3 separate ML models behind the scenes and takes **15 key inputs.**



🔗 **Live App URL:**
https://car-shipment-delay-prediction-8gxekgz52yqxspeqtu8svv.streamlit.app/

📁 **GitHub Repo:**
Amneetkaur24/car-shipment-delay-prediction

How This Project Drives Real Business Value

- **Smarter Operations Planning**
  By predicting delays and their financial impact, supply chain managers can **prioritize shipments**, reroute logistics, or increase buffer stock – **before** delays happen.
- **Cost Savings up to $450 per Vehicle**
  On average, the model can help **prevent $445–$665 in losses per delayed car**, by allowing early intervention in procurement and transport planning.
- **Improved Supplier Accountability**
  Supplier performance metrics and delay reasons now support **data-driven negotiations** and **SLA revisions**, especially for underperforming regions or routes.
- **Executive-Level Decision Support**
  The Power BI dashboard and Streamlit app offer an **interactive interface** for leadership to monitor, simulate, and act on delay trends instantly.

Future Work

Add Real-Time Data Integration
In future iterations, the model can integrate real-time data feeds (e.g., weather alerts, live supplier updates, strikes) to predict delays with greater accuracy and timeliness.

Enhance Feature Granularity
Current model uses part categories – future work can include **specific part codes** or **critical part tags** to precisely identify high-impact components like engine or transmission.

Expand Cost Analysis
Extend the financial impact model to include **indirect costs** (e.g., customer penalties, lost goodwill) and compare different **what-if scenarios** (e.g., fast shipping vs. delayed shipping costs).

Model Optimization
Continue tuning models (hyperparameter tuning, feature engineering) and test advanced methods like **XGBoost or SHAP** for even better performance and explainability.

Thank You!

🧑‍💼 **Project:** Car Shipment Delay Prediction

📍 **Focus:** Delay classification, delay days prediction, and financial impact estimation using machine learning

📈 **Goal:** Help manufacturing teams reduce losses and improve supply chain planning

🙏 **Thank you for your time!**
Feel free to ask about the models, app, or business insights – I'd love to walk through any part in more detail

Linkedin:- linkedin.com/in/amneetkaur24
Contact_Info:- amneet1224@gmail.com