PREDICTING VEHICLE SERVICE INTERVALS USING MACHINE LEARNING

CMPT-3830 – Machine Learning
Work Integrated Learning I – Fall 2025

Team members

Parminder Singh
Rajveer Singh
Jaswinder Singh
Farhan Mohammed
Manoj Bhatta

What we will be talking about-

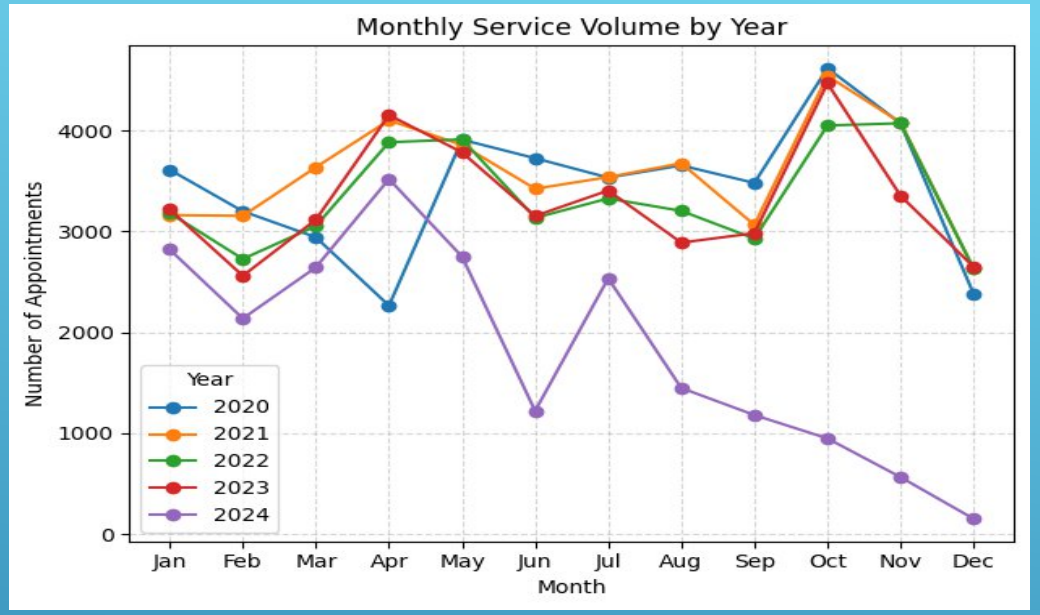
- Phase 1 highlights
- Client’s feedback for phase
- Improvements and insights after client’s feedback
- Power Bi dashboard
- ML application

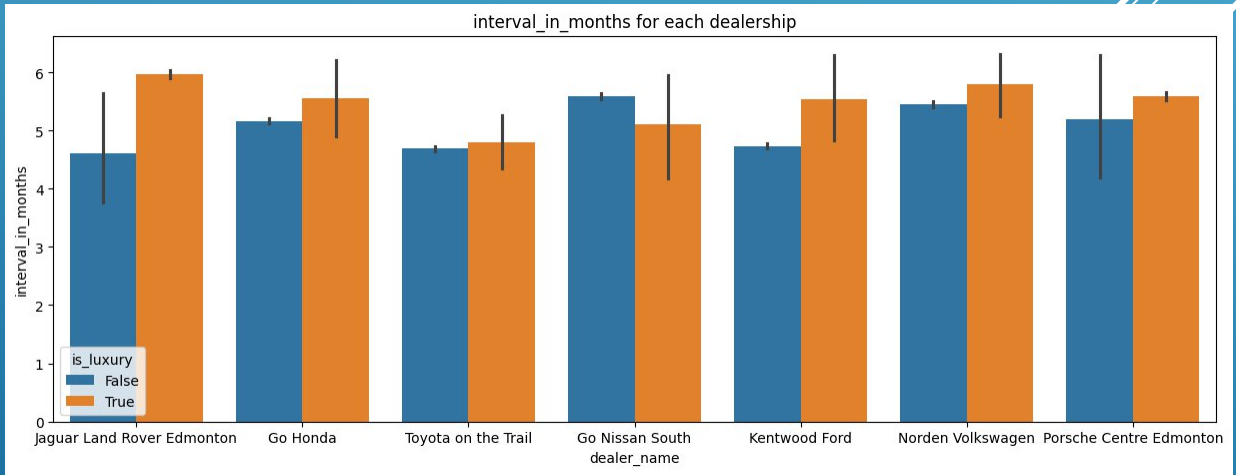
Client- Go Auto is a large, Canadian-owned family automotive group, one of the biggest in Canada, selling new/used vehicles, parts, and service across many brands (Ford, Mercedes, Nissan, etc.) through 70+ dealerships, offering a large inventory, 7-day service etc.

Problem Statement - analyzing Go Auto's service records to understand service frequency, dealership patterns, and differences between luxury and non-luxury vehicles.

Phase 1 highlights-

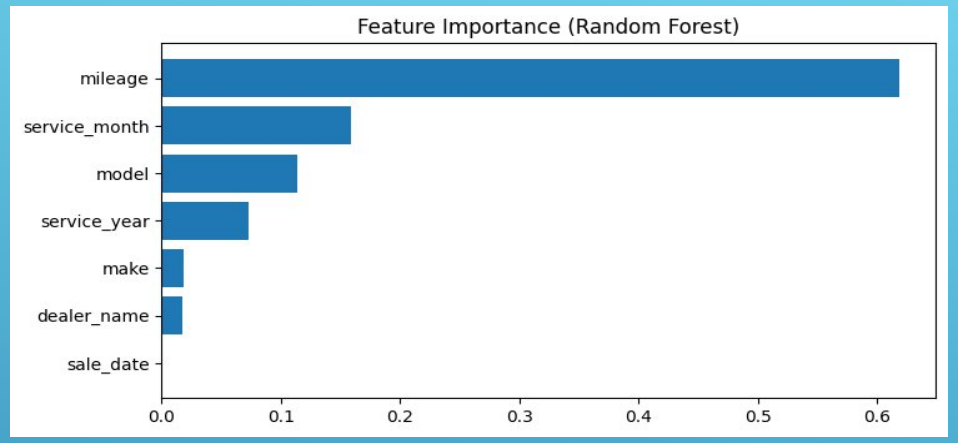
- Converted cost ranges to numeric midpoint values.
- Standardized make names using mapping dictionary.
- Dropped missing values and cleaned anomalies.
- Defined list of luxury brands.
- Created is_luxury boolean column.
- Enabled comparison between luxury and regular cars.
- Compared luxury vs regular service intervals.
- Analyzed dealership performance and service volume.
- Explored mileage class impact on service frequency.

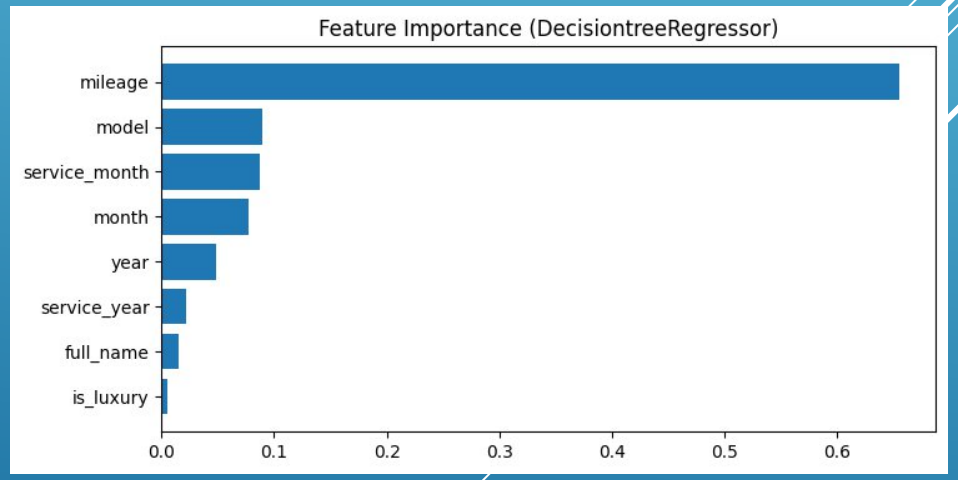




Phase 1 encoding and modeling

- Converted columns to numeric types.
- Dropped irrelevant columns for modeling.
- Label-encoded make, model, dealer_name.
- Defined X (features) and y (target: interval_in_months).
- Performed 80/20 train-test split.
- Models used: Linear Regression, Random Forest, Decision Tree, XGBoost.
- Evaluated using MSE, RMSE, R^2.
- Random Forest & DecisionTree performed best.



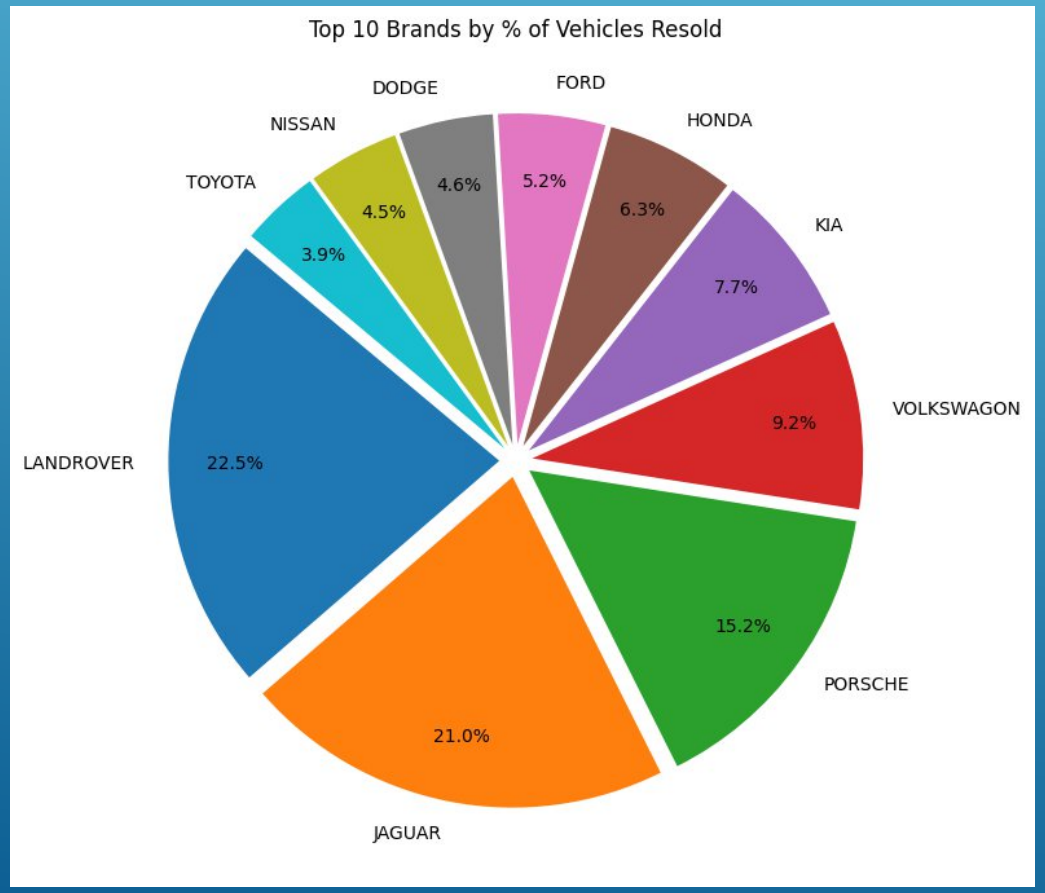


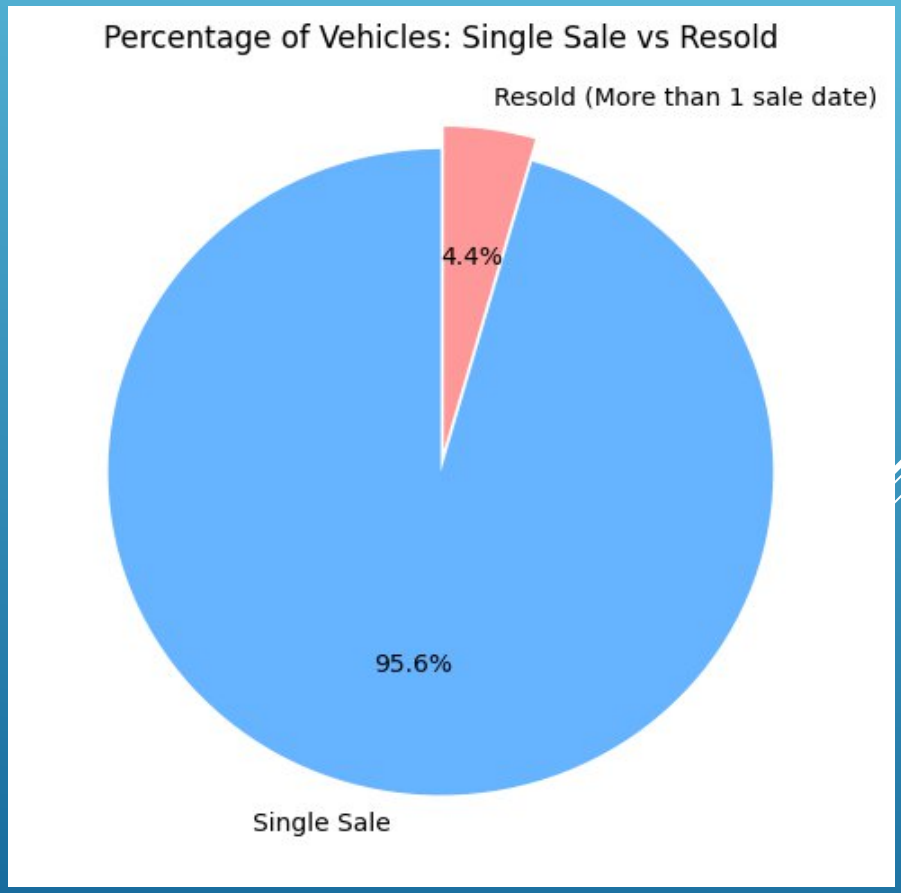
Client’s feedback:-

- Suggestion to built logic for determining if VIN is resold to another customer
- A logic for ON/OFF Make to check if same VIN brand comes to same brand dealership

Work and insights after client's feedback:-

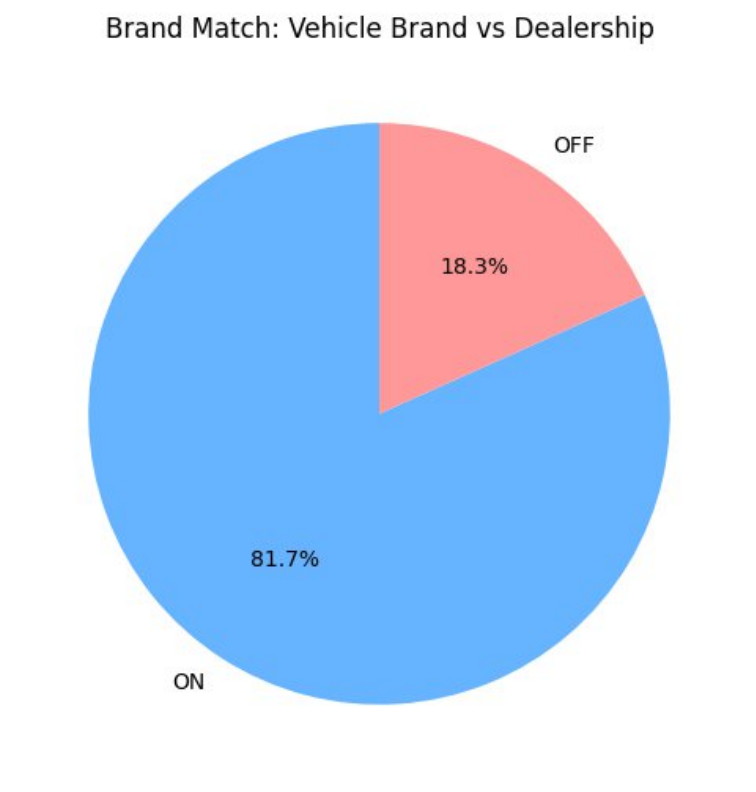
- Built a model to know if vin is resold and get to know almost 1747 VINs were resold.

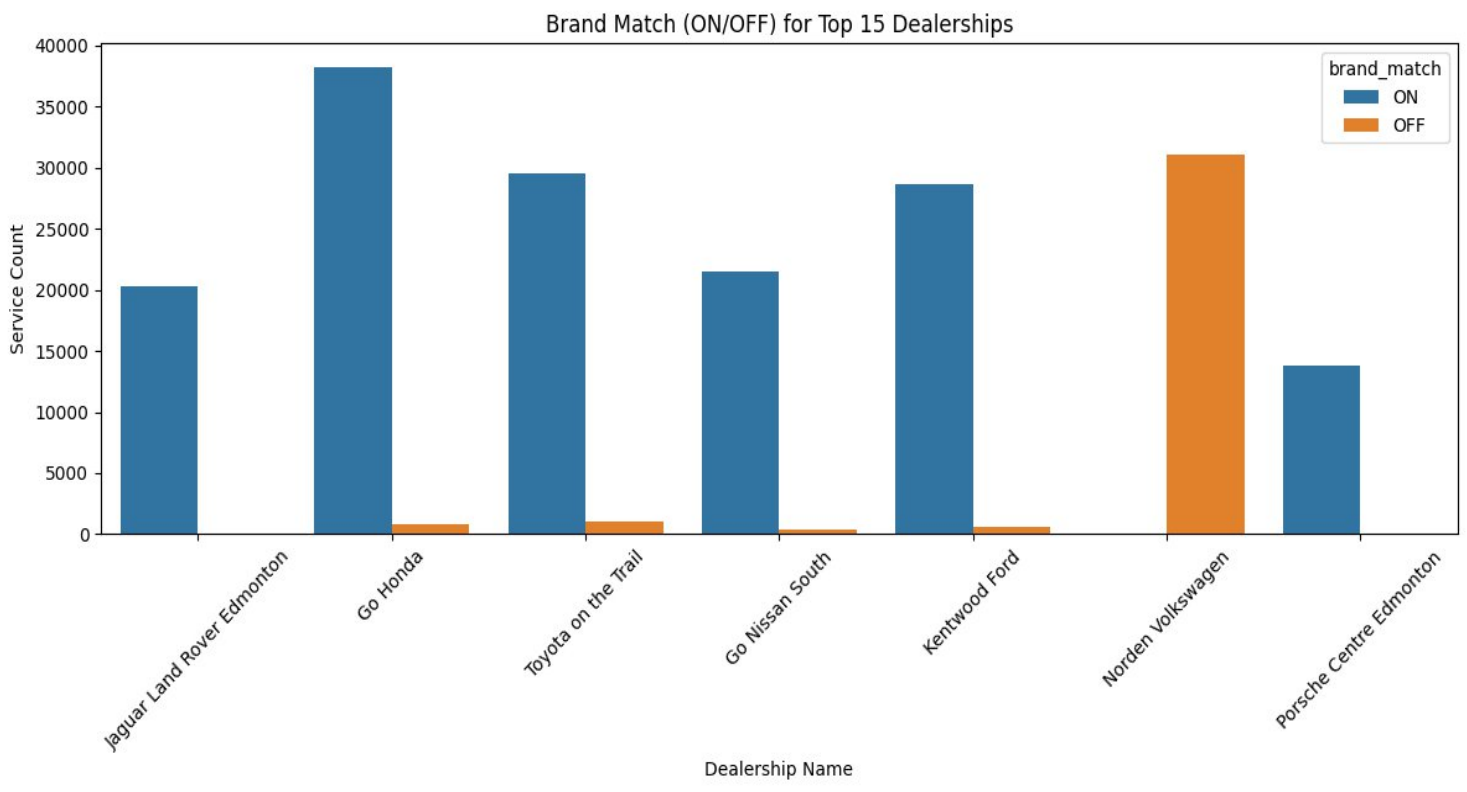




ON/OFF Make function and insights:-

- Built a function to check
- About 81 percent of vehicle go their brand dealership for service.





Phase 2:-

- Only used VINs which came back at least 1 time for service
- Finalized RandomForest model and trained it with necessary features.
- Built Streamlit application for prediction
- Built Powerbi dashboard

dealer_name

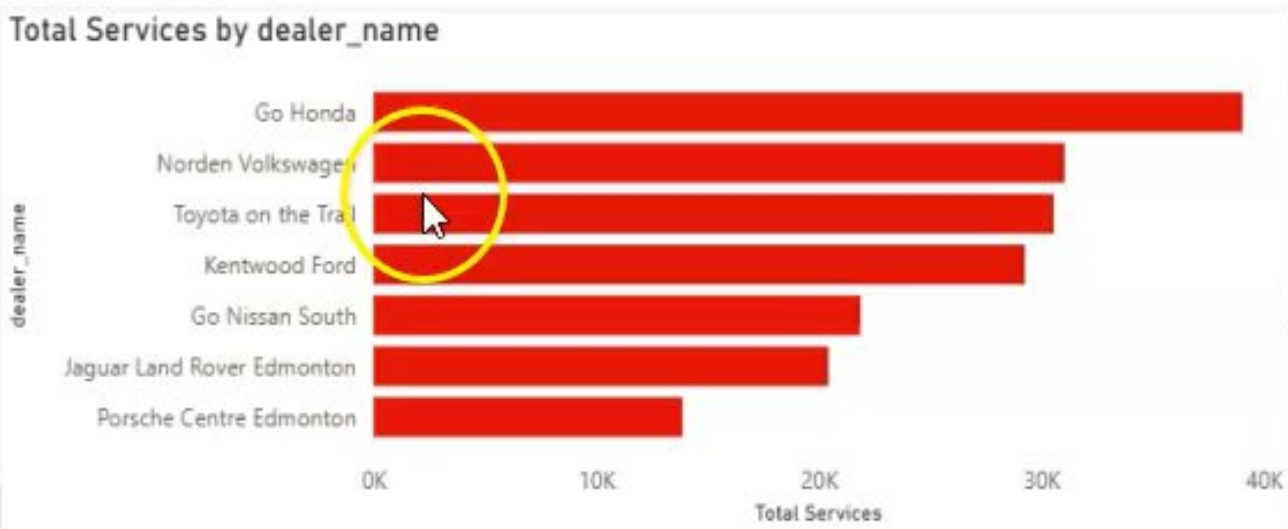
- [ ] Go Honda
- [ ] Go Nissan South
- [ ] Jaguar Land Rover Edmonton
- [ ] Kentwood Ford
- [ ] Norden Volkswagen
- [ ] Porsche Centre Edmonton
- [ ] Toyota on the Trail

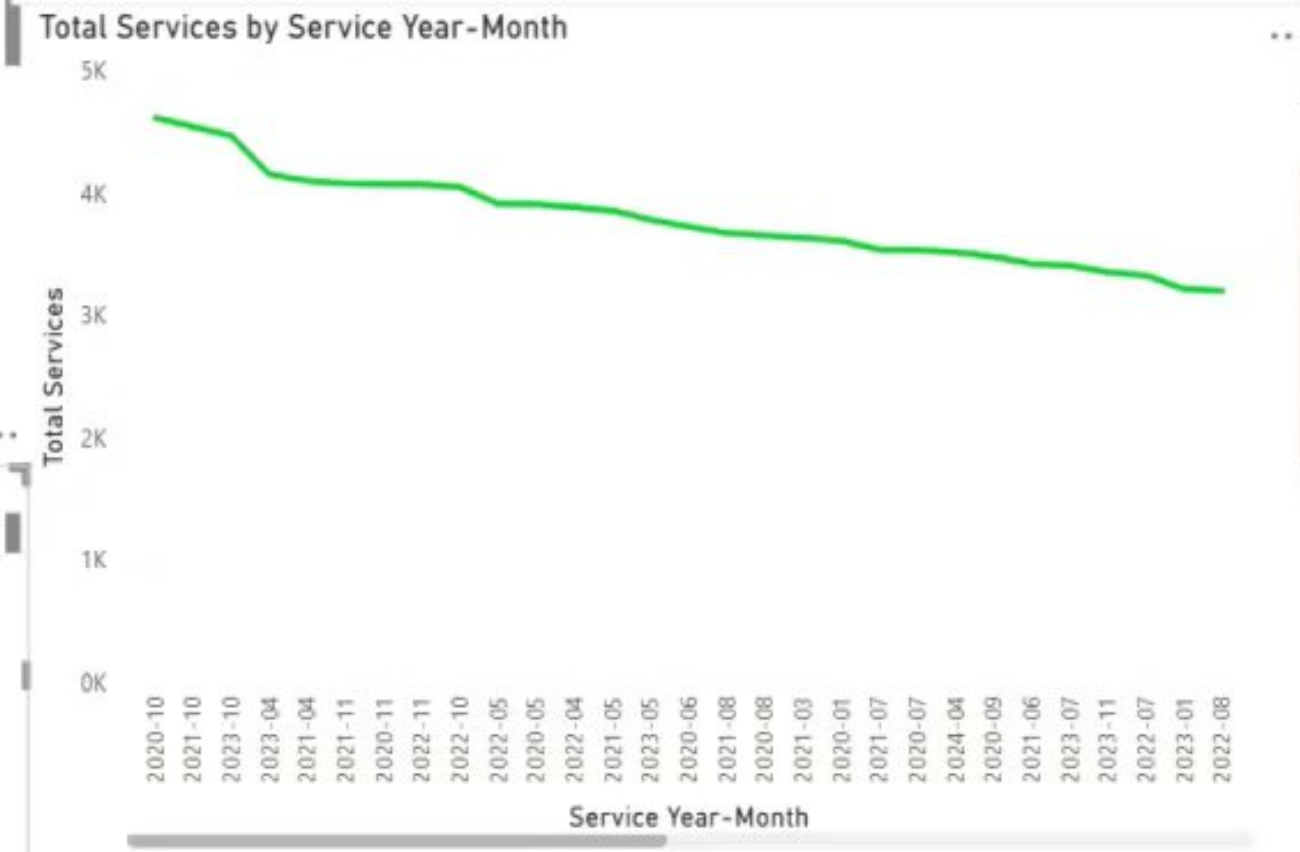
make

- [ ] ACURA
- [ ] ARCTIC MOTORS
- [ ] ASTONMARTIN
- [ ] AUDI
- [ ] BENTLY
- [ ] BMW
- [ ] BUICK
- [ ] CADILLAC
- [ ] CHEVORLET

Service Year-Month

- [ ] 2020-01
- [ ] 2020-02
- [ ] 2020-03
- [ ] 2020-04
- [ ] 2020-05
- [ ] 2020-06

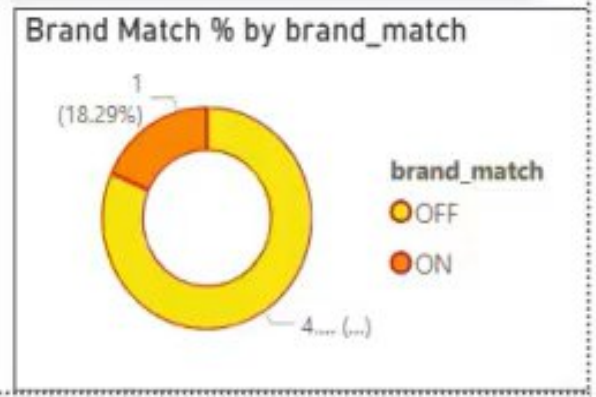




186K
Total Services

4.67
Median Interval (Months)

461.84
Average of cost



Navigation

Go to:

- Overview
- EDA
- Analysis
- Prediction

Go Auto Service Interval – Overview

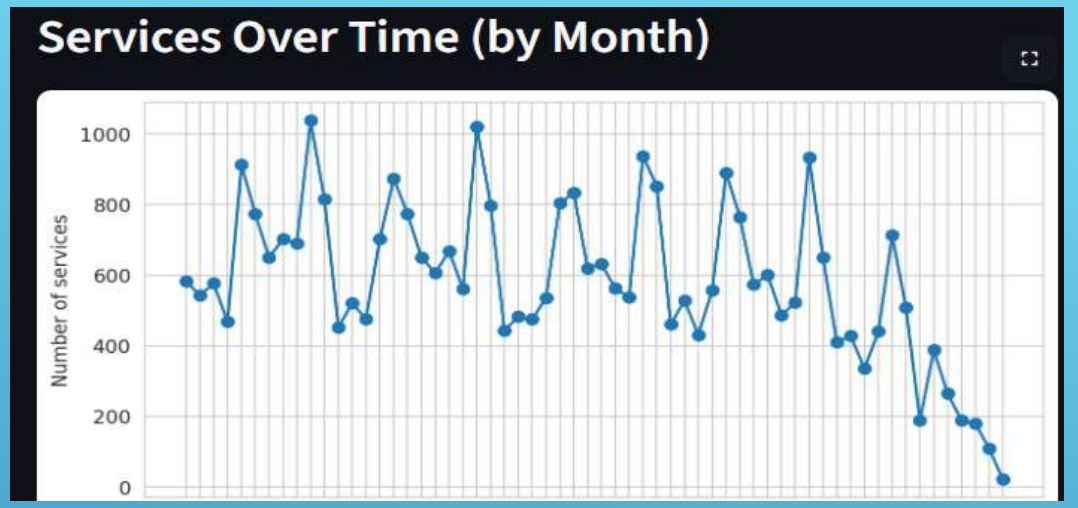
Total records	Unique VINs	Number of brands	Avg interval (months)
186,088	39,421	39	5.24

Sample of data used to train the model

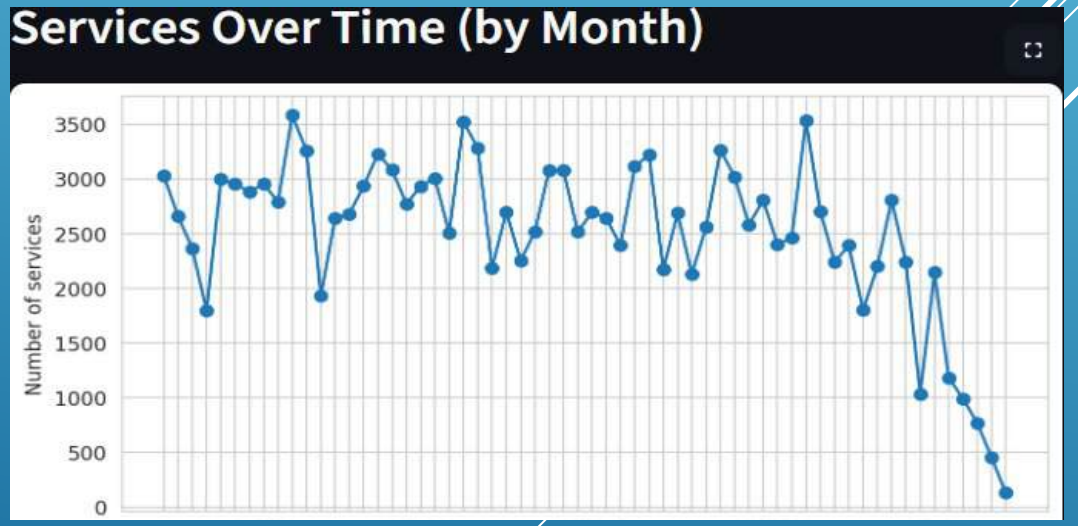
	service_date	make	vin	mileage	interval_in_months	is_luxury	dealer_name	cost
0	2021-11-30 00:00:00	LANDROVER	1000167976	14656	2.7333	☑	Jaguar Land Rover Edmonton	50
1	2022-02-20 00:00:00	LANDROVER	1000167976	18253	7.7333	☑	Jaguar Land Rover Edmonton	450
2	2022-10-10 00:00:00	LANDROVER	1000167976	33194	8.4333	☑	Jaguar Land Rover Edmonton	650
3	2023-05-29 00:00:00	HONDA	1000197869	9342	1.8333	☐	Go Honda	150
4	2023-07-23 00:00:00	HONDA	1000197869	15255	8.7333	☐	Go Honda	750
5	2024-04-10 00:00:00	HONDA	1000197869	26453	1.8667	☐	Go Honda	550
6	2024-06-05 00:00:00	HONDA	1000197869	29500	5.6667	☐	Go Honda	50
7	2021-06-05 00:00:00	TOYOTA	1000350530	73090	2.4	☐	Toyota on the Trail	150
8	2021-08-16 00:00:00	TOYOTA	1000350530	74322	9.5	☐	Go Nissan South	50
9	2023-10-03 00:00:00	TOYOTA	1000402957	7209	0.4667	☐	Toyota on the Trail	

Conclusion

- Overall dealerships have more regular vehicles coming for service.
- Less number of luxury vehicle services is cause of less ownership and different drive behavior than regular cars.
- Mileage is dominant factor for car owners to return for service if no other issue.
- Still both luxury and regular vehicles comes around same time period.



Luxury vehicle graph



Regular car service graph

Reflection

- Mostly return time is based on mileage and cost can be somewhat factor for some dealerships.
- Feedback helped us improve and get some new insights.
- In future, with more detail like service type can be helpful to get better prediction for return time of customer.
- Exploring more about each customer driving pattern can be helpful for future improvements.

Thank you Go Auto

For giving us opportunity to get familiar with real life experience of data science world problems and motivating us to improve.