US USED CARS DATASET

NAME:

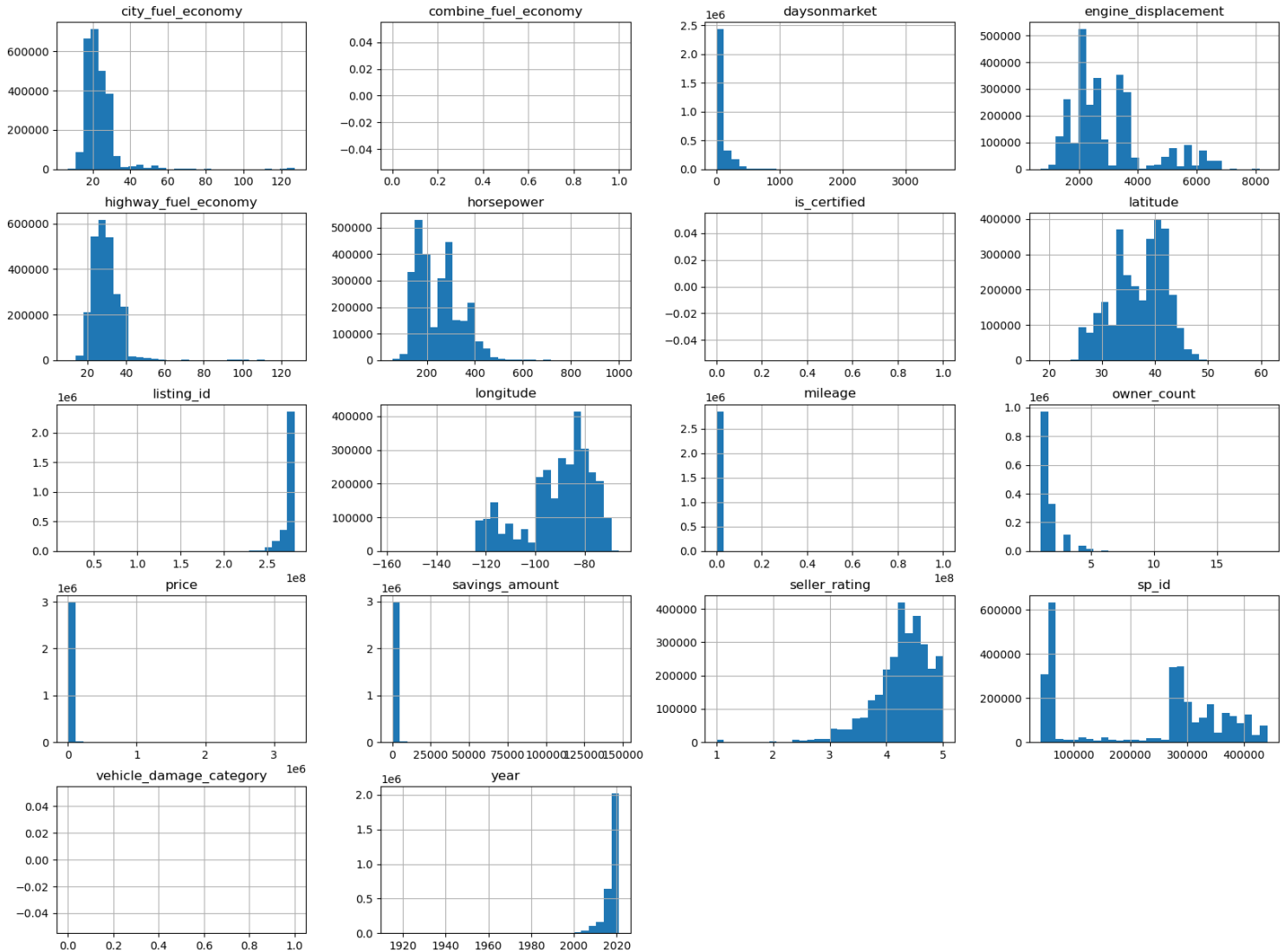
HIEU HUA

VUONG PHAT

HUSAN IFTAKHAR

Introduction:

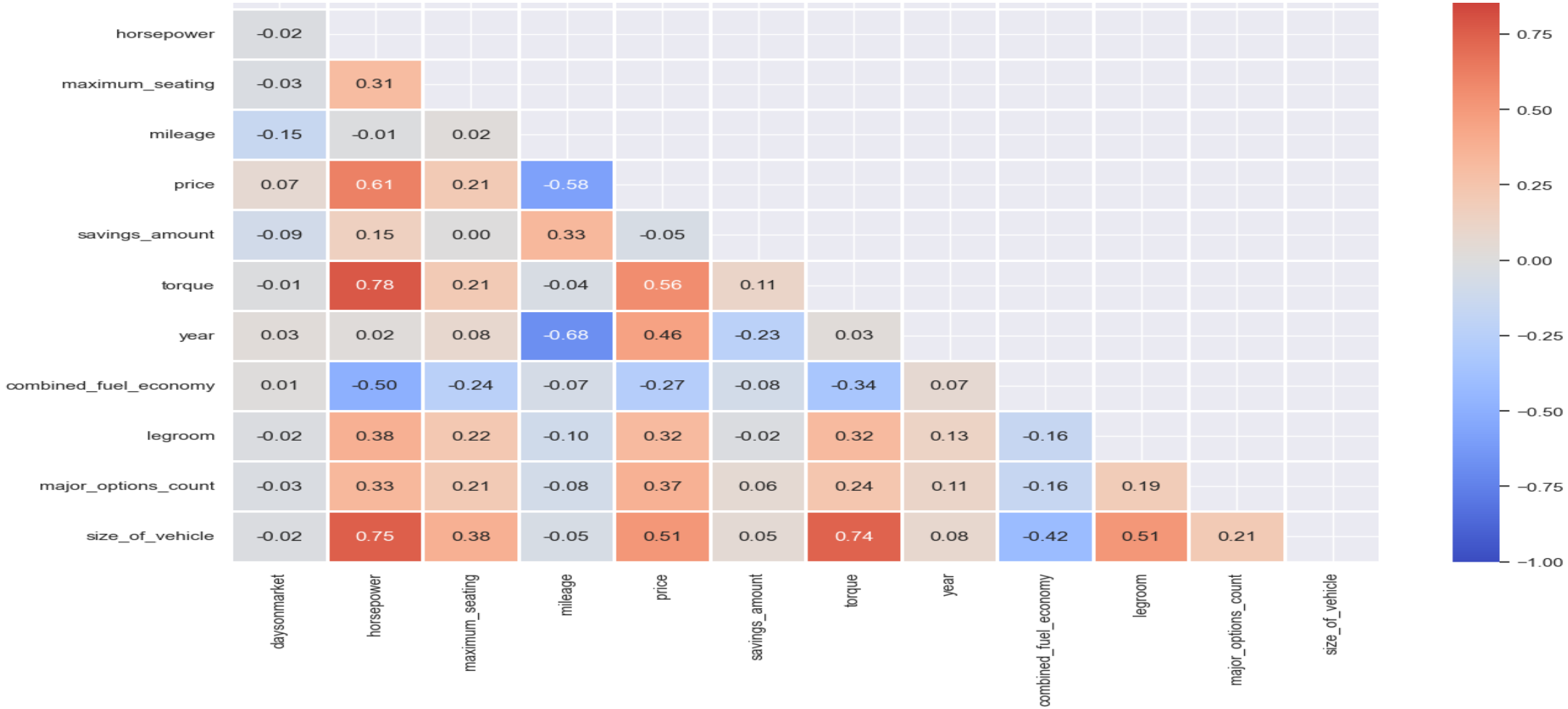
- **Business Objectives**: Optimize pricing strategies, understand vehicle demand, enhance customer decision-making, and improve inventory management.
- **Initial Analysis Goals**: Identify pricing factors, predict time on market, and analysis
- **Success Metrics**: RMSE, MAE, R-squared for models.
- **Technologies**: Python (Pandas, NumPy, Scikit-Learn), SQL, Matplotlib, Seaborn.

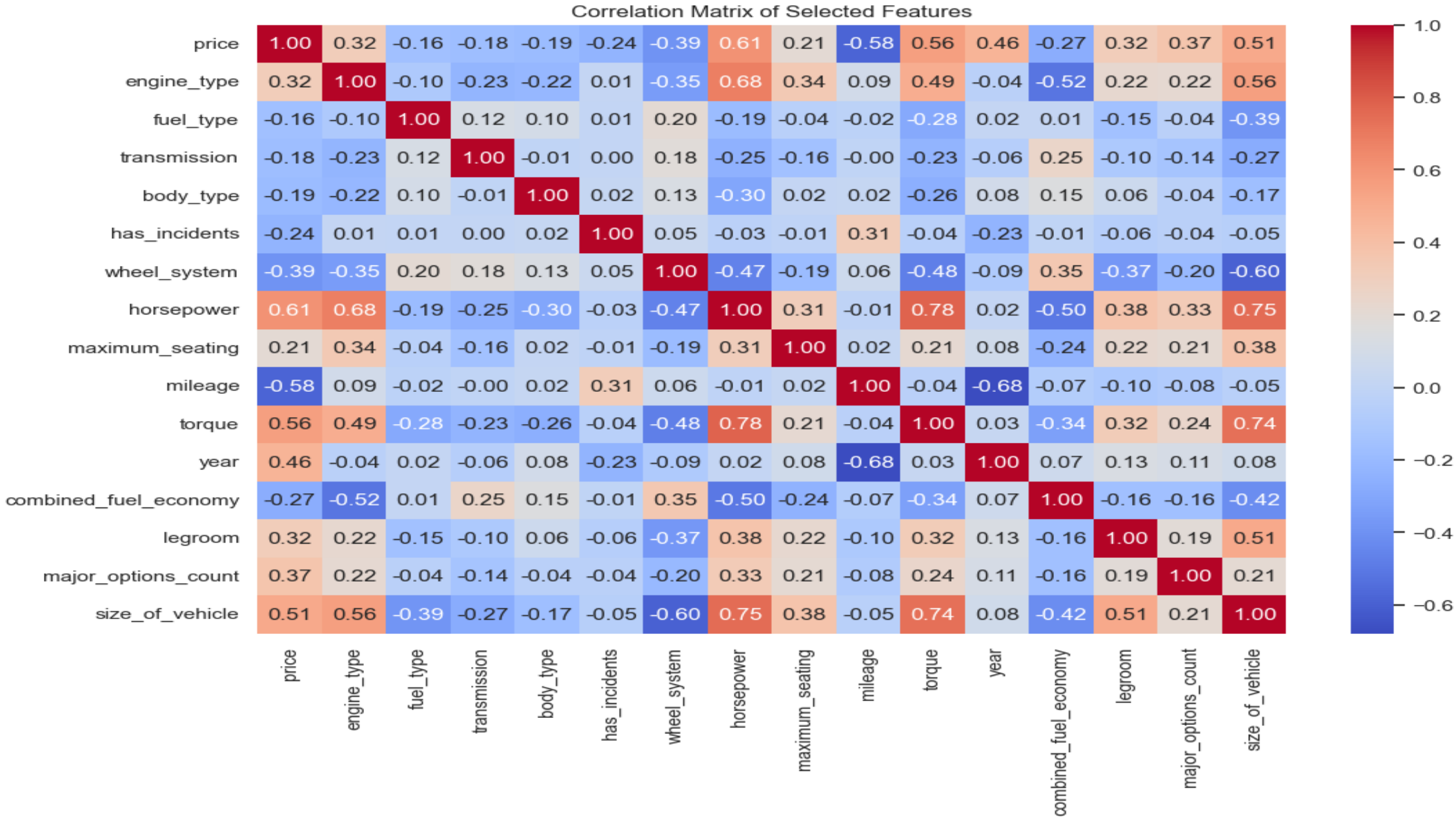

city_fuel_economy
combine_fuel_economy
daysonmarket
1e6
engine_displacement
highway_fuel_economy
horsepower
is_certified
latitude
listing_id
1e6
1e8
longitude
mileage
1e6
1e8
owner_count
1e6
price
1e6
savings_amount
1e6
seller_rating
sp_id
vehicle_damage_category
1e6
year
1e6

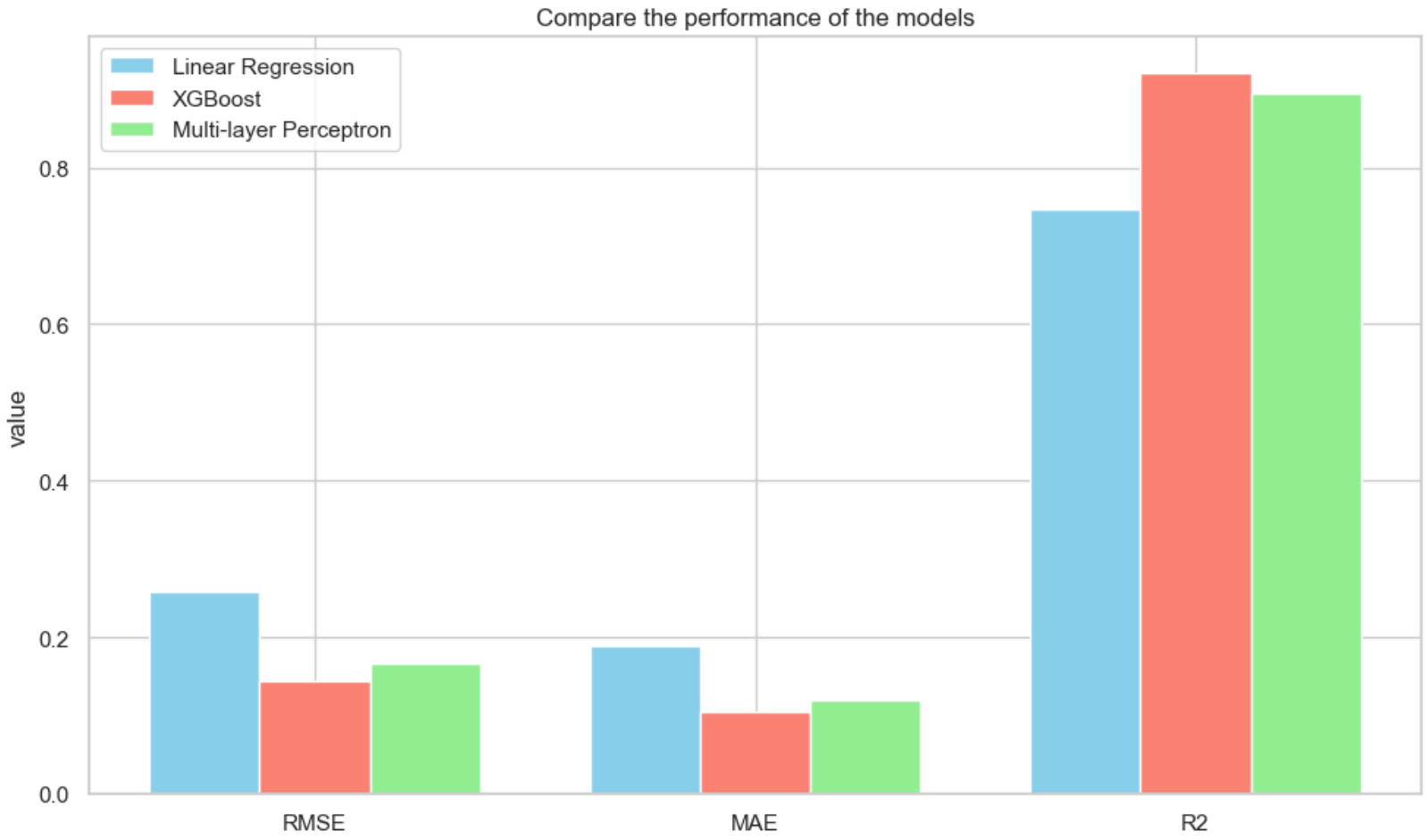
Phase 2

Phase 3

- The average listing duration is 75 days
- The average speed is 243 horsepower
- The average vehicle is 5 seats
- Most of vehicles being sold are relatively new, with an average mileage of 20,835 miles.
- The average price of a vehicle is $29933.37
- The average savings is $554.62.
- …….


horsepower -0.02
maximum_seating -0.03 0.31
mileage -0.15 -0.01 0.02
price 0.07 0.61 0.21 -0.58
savings_amount -0.09 0.15 0.00 0.33 -0.05
torque -0.01 0.78 0.21 -0.04 0.56 0.11
year 0.03 0.02 0.08 -0.68 0.46 -0.23 0.03
combined_fuel_economy 0.01 -0.50 -0.24 -0.07 -0.27 -0.08 -0.34 0.07
legroom -0.02 0.38 0.22 -0.10 0.32 -0.02 0.32 0.13 -0.16
major_options_count -0.03 0.33 0.21 -0.08 0.37 0.06 0.24 0.11 -0.16 0.19
size_of_vehicle -0.02 0.75 0.38 -0.05 0.51 0.05 0.74 0.08 -0.42 0.51 0.21
daysonmarket
horsepower
maximum_seating
mileage
price
savings_amount
torque
year
combined_fuel_economy
legroom
major_options_count
size_of_vehicle
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

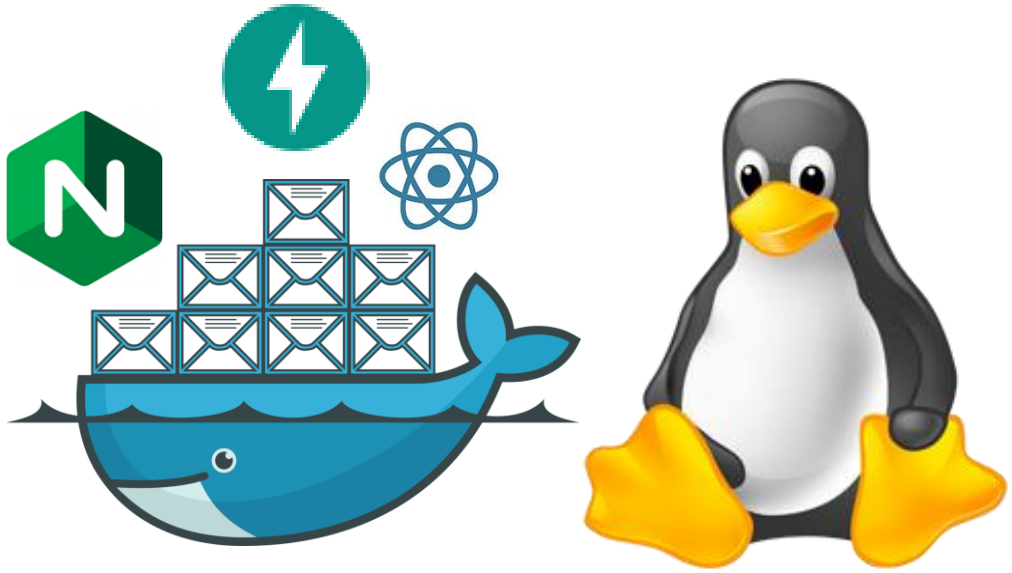

Correlation Matrix of Selected Features
price 1.00 0.32 -0.16 -0.18 -0.19 -0.24 -0.39 0.61 0.21 -0.58 0.56 0.46 -0.27 0.32 0.37 0.51
engine_type 0.32 1.00 -0.10 -0.23 -0.22 0.01 -0.35 0.68 0.34 0.09 0.49 -0.04 -0.52 0.22 0.22 0.56
fuel_type -0.16 -0.10 1.00 0.12 0.10 0.01 0.20 -0.19 -0.04 -0.02 -0.28 0.02 0.01 -0.15 -0.04 -0.39
transmission -0.18 -0.23 0.12 1.00 -0.01 0.00 0.18 -0.25 -0.16 -0.00 -0.23 -0.06 0.25 -0.10 -0.14 -0.27
body_type -0.19 -0.22 0.10 -0.01 1.00 0.02 0.13 -0.30 0.02 0.02 -0.26 0.08 0.15 0.06 -0.04 -0.17
has_incidents -0.24 0.01 0.01 0.00 0.02 1.00 0.05 -0.03 -0.01 0.31 -0.04 -0.23 -0.01 -0.06 -0.04 -0.05
wheel_system -0.39 -0.35 0.20 0.18 0.13 0.05 1.00 -0.47 -0.19 0.06 -0.48 -0.09 0.35 -0.37 -0.20 -0.60
horsepower 0.61 0.68 -0.19 -0.25 -0.30 -0.03 -0.47 1.00 0.31 -0.01 0.78 0.02 -0.50 0.38 0.33 0.75
maximum_seating 0.21 0.34 -0.04 -0.16 0.02 -0.01 -0.19 0.31 1.00 0.02 0.21 0.08 -0.24 0.22 0.21 0.38
mileage -0.58 0.09 -0.02 -0.00 0.02 0.31 0.06 -0.01 0.02 1.00 -0.04 -0.68 -0.07 -0.10 -0.08 -0.05
torque 0.56 0.49 -0.28 -0.23 -0.26 -0.04 -0.48 0.78 0.21 -0.04 1.00 0.03 -0.34 0.32 0.24 0.74
year 0.46 -0.04 0.02 -0.06 0.08 -0.23 -0.09 0.02 0.08 -0.68 0.03 1.00 0.07 0.13 0.11 0.08
combined_fuel_economy -0.27 -0.52 0.01 0.25 0.15 -0.01 0.35 -0.50 -0.24 -0.07 -0.34 0.07 1.00 -0.16 -0.16 -0.42
legroom 0.32 0.22 -0.15 -0.10 0.06 -0.06 -0.37 0.38 0.22 -0.10 0.32 0.13 -0.16 1.00 0.19 0.51
major_options_count 0.37 0.22 -0.04 -0.14 -0.04 -0.04 -0.20 0.33 0.21 -0.08 0.24 0.11 -0.16 0.19 1.00 0.21
size_of_vehicle 0.51 0.56 -0.39 -0.27 -0.17 -0.05 -0.60 0.75 0.38 -0.05 0.74 0.08 -0.42 0.51 0.21 1.00
price
engine_type
fuel_type
transmission
body_type
has_incidents
wheel_system
horsepower
maximum_seating
mileage
torque
year
combined_fuel_economy
legroom
major_options_count
size_of_vehicle
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6



Phase 4 & 5 : Modeling & Evaluation

Phase 6 Overview: Deployment and Features

1. FastAPI backend serves HTTP endpoints
2. React UI communicates with FastAPI
3. Current capabilities:
   - Predict car prices
   - Add new data rows
   - Retrain models
4. Docker-based deployment with one command docker compose up –d
5. Logs can be seen using docker logs command



Add data and retrain

Training Row Addition form

body_type	SUV / Crossover
engine_type	I4
fuel_type	Gasoline
has_incidents	False
transmission	A
wheel_system	AWD
horsepower	177
maximum_seating	5
mileage	7
torque	200
year	2019
combined_fuel_economy	25
legroom	76.3
major_options_count	1
size_of_vehicle	426.6
price	10.532

Add Training Row

Retrain

Get predictions

Price prediction form

body_type	SUV / Crossover
engine_type	I4
fuel_type	Gasoline
has_incidents	False
transmission	A
wheel_system	AWD
horsepower	177
maximum_seating	5
mileage	7
torque	200
year	2019
combined_fuel_economy	25
legroom	76.3
major_options_count	1
size_of_vehicle	426.6

Get Prediction

Prediction Results

Linear Regression: 10.139370063164234

XGBoost: 9.092524528503418

MLP: 10.1796875

Close

POST	/api/v1/predict/price	Get Price Predictions
GET	/api/v1/predict/label_mappings	Get Label Mappings
POST	/api/v1/predict/insert_row	Add Training Row
POST	/api/v1/predict/__retrain_models__	Retrain Models
GET	/	Redirect To Docs

Deployment plans

1. Linux machine on cloud with docker
2. Google Cloud IDX

Future plans

- Making UI look good
- Testing
- Monitoring
- Data backup
- Adding more recommendation models for other useful columns like **days_on_market**
- Manually verifying the newly trained model's accuracy