



CP Project

Done by: Shaimerdin Shyngyskhan, Marat Nurzhan, Taubai Zangar

Basic overview and “pain”

*Which variables are significant in predicting the price of a car.

*How well those variables describe the price of a car.

*Significant increase for car price in the Kazakhstan's market, may put people in difficult situation in terms of when and for what type of car there would be a decent price.

«Надежды уже нет». Цены на машины с пробегом рванули вверх (таблица)



What have been done before?

Our team in general use the following materials for the project because it provides, in our opinion, basic and broad guide for us.



Predicting Used Car Prices with Machine Learning Techniques

Comparing Performance of Five Different ML Models

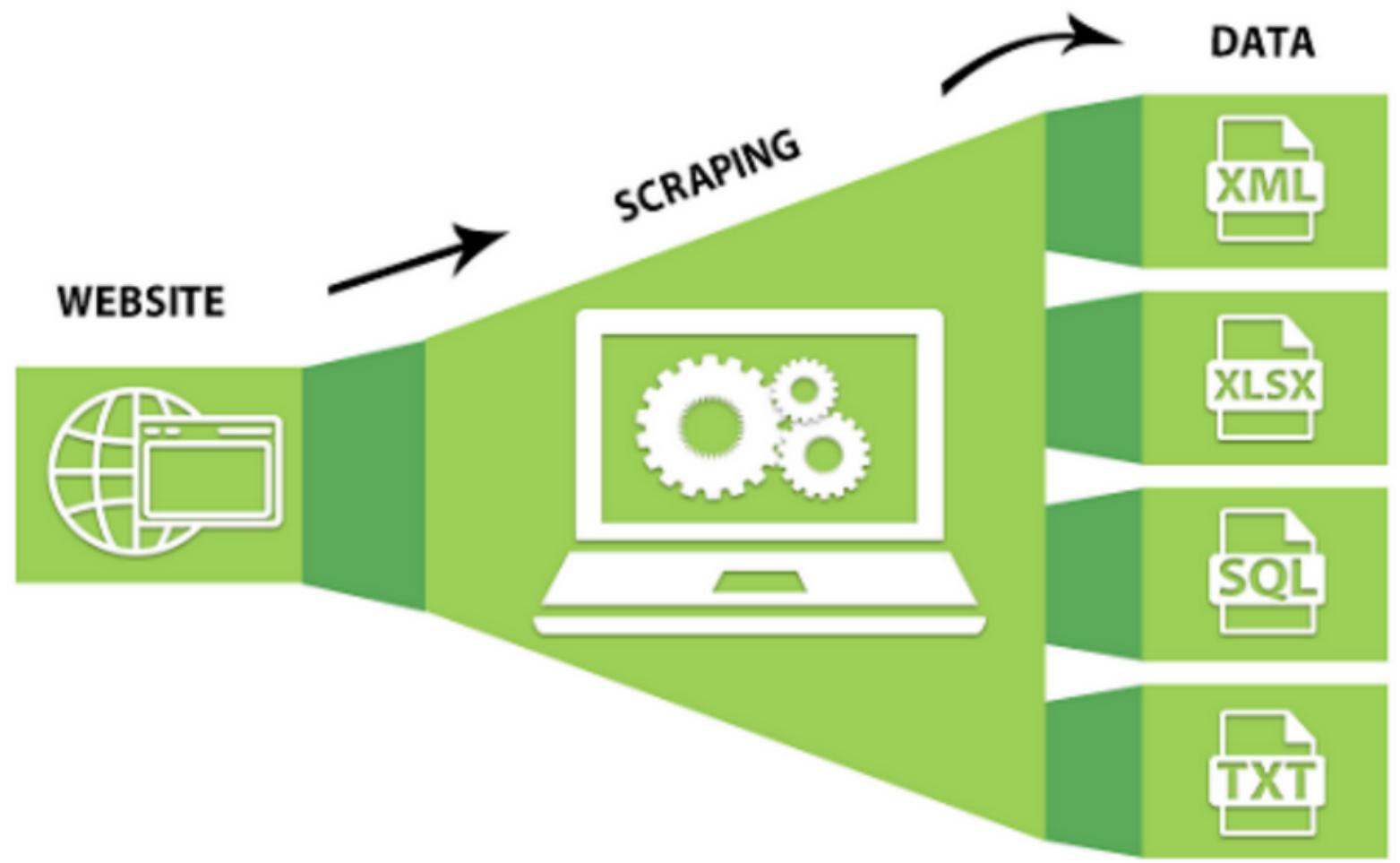
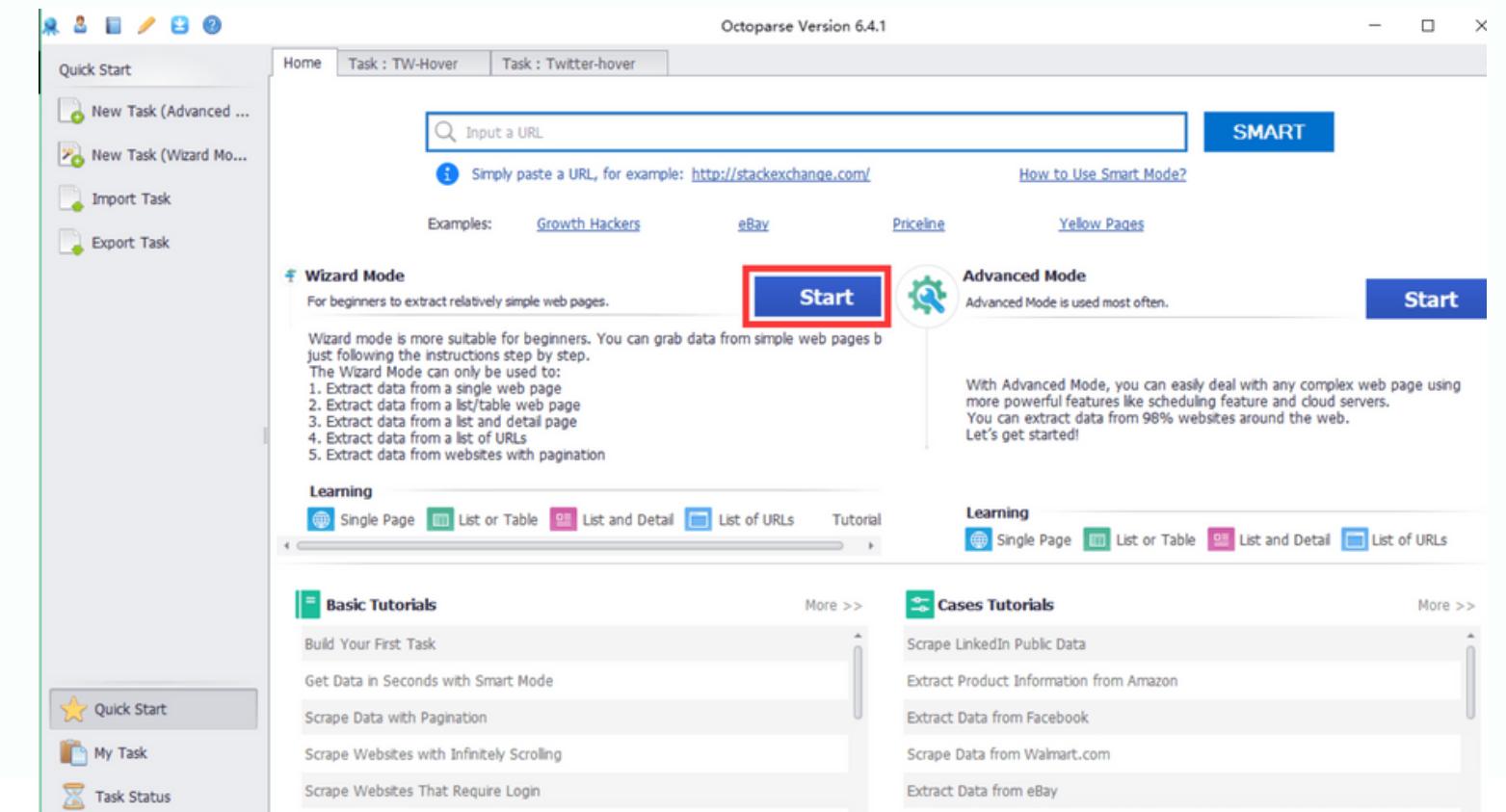
Medium / Enes Gokce / Jan 10, 2020

Link to the source materials:

<https://towardsdatascience.com/predicting-used-car-prices-with-machine-learning-techniques-8a9d8313952>

Source for the dataset

- 1) SCRAPING
- 2) PARSING
- 3) WE USED OCTOPARSE



Parameters, according to which we will attempt to do a prediction

- Fuel type
- Price
- Gearbox type
- Body type
- Mileage
- Color
- Engine size
- City
- Mark
- Model

The screenshot shows a car search interface with the following features:

- Where to search:** Options include Где искать (Where to search), Алматы (Almaty), Нур-Султан (Астана) (Nur-Sultan (Astana)), Шымкент (Shymkent), Актобе (Aktobe), Караганда (Karaganda), Тараз (Taz), and еще (more).
- Brand:** Options include Марка (Brand), ВАЗ (Lada), Toyota, Mercedes-Benz, Volkswagen, Nissan, Audi, and еще (more).
- Model:** Options include Модель (Model), Camry, Land Cruiser Prado, Land Cruiser, Corolla, RAV 4, Highlander, and еще (more).
- Status:** Options include Все (All), Новая (New), С пробегом (With mileage), С фото (With photo), Растаможен (Customs cleared), and Аварийная/Не на ходу (Emergency/Not running).
- Advanced Search:** A large yellow button labeled Расширенный поиск (Advanced search) is visible.
- Filters:** Various dropdown menus and input fields for filtering search results:
 - Страна происхождения ... (Country of origin ...)
 - Кузов (Body type)
 - Двигатель (Engine)
 - КПП (Transmission)
 - Расположение руля (Steering wheel location)
 - Привод (Drive type)
 - Пробег не более, км (Mileage not more than, km)
 - Наличие (Availability)
 - Объем двигателя, л (Engine volume, l)
 - Цвет (Color)
 - Не важно (It doesn't matter)
 - В наличии (In stock)
 - На заказ (On order)
 - от (from)
 - до (to)
- Other:** A checkbox for металлик (Metallic) and a small image of a car.

Parsing process

The Description was in one field

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	listregion	Title	Price	Description											
2	Алматы	Renault L	7 207 680	2020 г., новый седан, 1.6 л, бензин, КПП автомат, синий, металлик											
3	Алматы	Toyota Av	4 200 000	2003 г., Б/у седан, 2 л, бензин, КПП автомат, с пробегом 200 000 км, серебристый, металлик, МАШИНА ПОЛНОСТЬЮ В РОДНОЙ...											
4	Алматы	Renault Sa	7 056 000	2020 г., новый хэтчбек, 1.6 л, бензин, КПП автомат, коричневый											
5	Алматы	Mitsubish	12 500 00	2006 г., Б/у микроавтобус, 3 л, бензин, правый руль, КПП автомат, с пробегом 180 000 км, серебристый, металлик, люк, в...											
6	Алматы	Infiniti QX	28 050 00	2018 г., Б/у внедорожник, 5.6 л, бензин, КПП автомат, с пробегом 22 000 км, белый, Продажа от официального дилера Infi...											
7	Алматы	Toyota Co	4 380 000	2011 г., Б/у седан, 1.6 л, бензин, КПП автомат, белый, Продажа от официального дилера Infiniti Автомир. По тех паспорт...											
8	Алматы	Nissan Pa	11 500 00	2012 г., Б/у внедорожник, 5.6 л, бензин, КПП автомат, с пробегом 105 000 км, белый, Продажа от официального дилера Inf...											
9	Алматы	Lexus LX	5 57 000 00	2020 г., новый внедорожник, 5.7 л, бензин, КПП автомат, черный, Black Vision II +, имеются: доводчики дверей, тонирова...											
10	Алматы	BA3 (Lada)	4 800 000	2019 г., Б/у минивэн, 1.6 л, бензин, КПП механика, с пробегом 90 000 км, черный, литые диски, тонировка, ветровики, ре...											
11	Алматы	Hyundai A4	550 000	2013 г., Б/у хэтчбек, 1.6 л, бензин, КПП автомат, красный, металлик, литые диски, тонировка, ветровики, противотуманки...											
12	Алматы	Volkswage	1 750 000	1995 г., Б/у хэтчбек, 2 л, бензин, КПП механика, с пробегом 120 000 км, синий, металлик, литые диски, тонировка, ветро...											
13	Алматы	Toyota Co	3 700 000	2012 г., Б/у седан, 1.4 л, бензин, КПП механика, с пробегом 90 000 км, серебристый, металлик, ветровики, хрустальная о...											
14	Алматы	Toyota La	9 150 000	2007 г., Б/у внедорожник, 2.7 л, газ-бензин, КПП автомат, с пробегом 196 000 км, черный, литые диски, тонировка, люк,...											
15	Алматы	Mercedes-	26 500 00	2019 г., Б/у кроссовер, 2 л, бензин, КПП автомат, с пробегом 4 000 км, синий, литые диски, тонировка, ветровики, обогр...											
16	Алматы	Lexus IS	2 10 900 00	2014 г., Б/у седан, 2.5 л, бензин, КПП типтроник, красный, металлик, литые диски, тонировка, обвес, обогрев зеркал, хр...											
17	Алматы	Hyundai E	5 800 000	2015 г., Б/у седан, 1.6 л, бензин, КПП типтроник, серый, литые диски, тонировка, ветровики, противотуманки, корректор...											
18	Алматы	Mazda Cr	1 250 000	1994 г., Б/у седан, 2 л, бензин, КПП автомат, вишня, литые диски, тонировка, багажник, обогрев зеркал, велюр, аудиосис...											
19	Алматы	Toyota Ve	8 900 000	2010 г., Б/у кроссовер, 2.7 л, бензин, КПП автомат, с пробегом 146 000 км, серый, металлик, литые диски, тонировка, ве...											
20	Алматы	BA3 (Lada)	3 200 000	2015 г., Б/у седан, 1.6 л, бензин, КПП механика, с пробегом 18 000 км, серебристый, велюр, аудиосистема, продам lada G...											
21	Алматы	Lexus RX	35 400 000	2002 г., Б/у кроссовер, 3 л, бензин, КПП автомат, белый, металлик, люк, ветровики, фаркоп, рейлинги, литые диски, бага...											
22	Алматы	Mercedes-	44 500 00	2018 г., Б/у седан, 4 л, бензин, КПП автомат, с пробегом 34 000 км, Покупалась у официального дилера. Amg +. Состояние...											
23	Алматы	Cadillac X	32 500 00	2020 г., новый кроссовер, 2 л, бензин, КПП автомат											
24	Алматы	Toyota La	17 500 00	2008 г., Б/у внедорожник, 4 л, бензин, КПП автомат, с пробегом 118 000 км, серый, САМАЯ ПОЛНАЯ КОМПЛЕКТАЦИЯ! Есть люк,...											
25	Алматы	BA3 (Lada)	2 000 000	2012 г., Б/у седан, 1.6 л, бензин, КПП механика, белый, СРОЧНО сейчас нужно забрать состояния идеал родной краска 100%...											
26	Алматы	Renault Sa	6 679 680	2020 г., новый хэтчбек, 1.6 л, бензин, КПП механика, серый, металлик											
27	Алматы	Renault Lc	7 207 680	2020 г., новый седан, 1.6 л, бензин, КПП механика											



Kia Morning

2011 г., Б/у хэтчбек, 1 л, бензин, КПП автомат, с пробегом 128 000 км, белый, металлик, тонировка, литые диски, багажн...

3 200 000 ₮

В кредит

83 219 ₮/мес



Kia Rio

2015 г., Б/у хэтчбек, 1.4 л, бензин, КПП автомат, коричневый, металлик

4 970 000 ₮

В кредит

145 405 ₮/мес



Geely Emgrand EC7

2012 г., Б/у седан, 1.5 л, бензин, КПП механика, с пробегом 220 000 км, серебристый, аудиосистема, MP3, DVD, сабвуфер,...

1 900 000 ₮

В кредит

36 309 ₮/мес

Parsing process

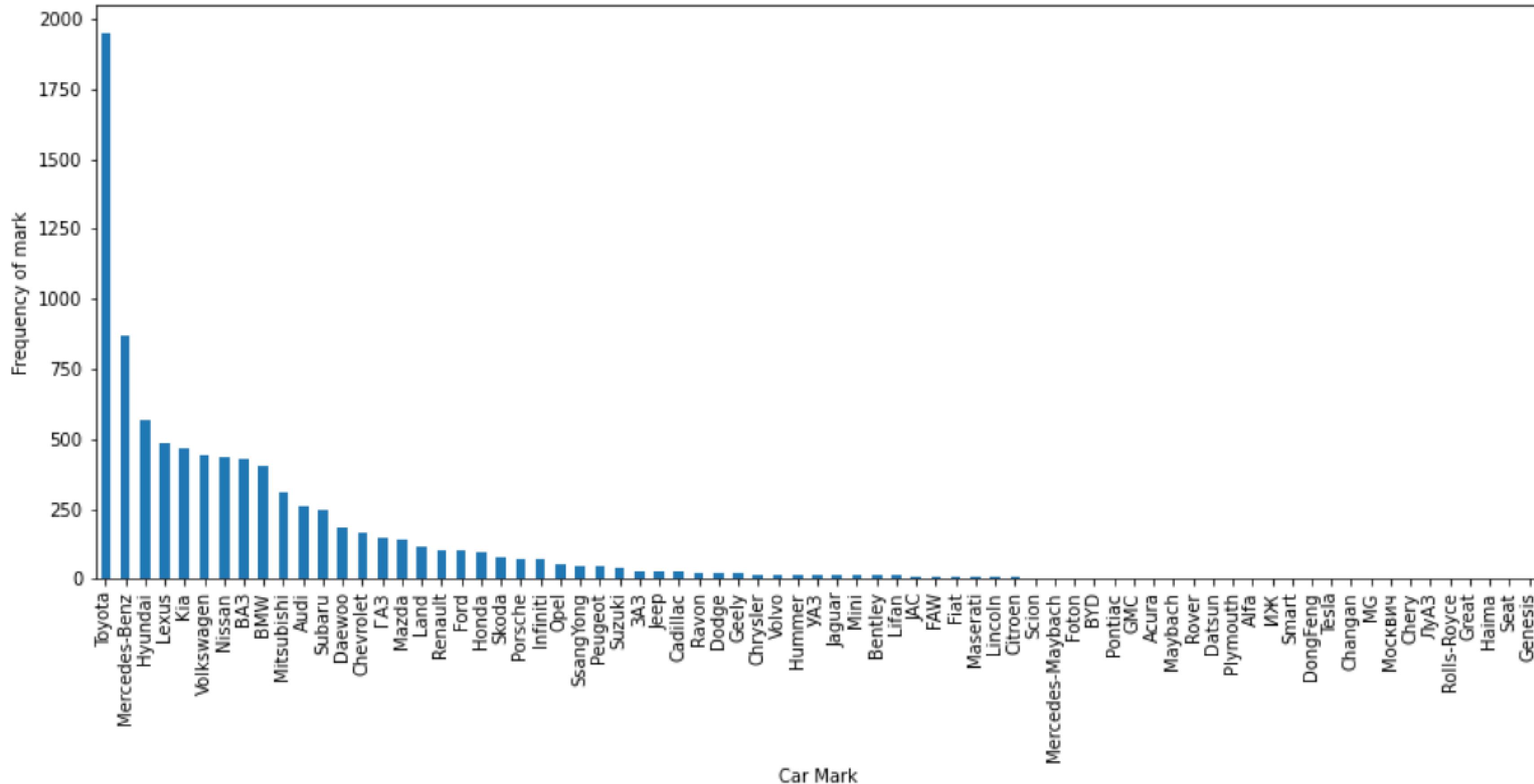
Divided by fields and created a normal database

The screenshot shows a web interface for a car marketplace. At the top, there's a navigation bar with links like 'Машины', 'Запчасти', 'Ремонт и услуги', 'Коммерческие', 'Прочее', 'Почитать', 'Kolesa Гид'. Below the navigation, a message says 'Run Completed!' with a green checkmark icon. It provides details about the task: 'Купить автомобиль в Алматы. Продажа машин в А...', 'Run time: 1h 30min 56s', and 'Data extracted: 20000 line(s) (1164 duplicates)'. There are two buttons at the bottom right: 'Export Later' and 'Export Data'. At the bottom, there's a footer with links: '< 1 ... 1998 1999 2000 >', 'Cloud Backup', and a small square icon.

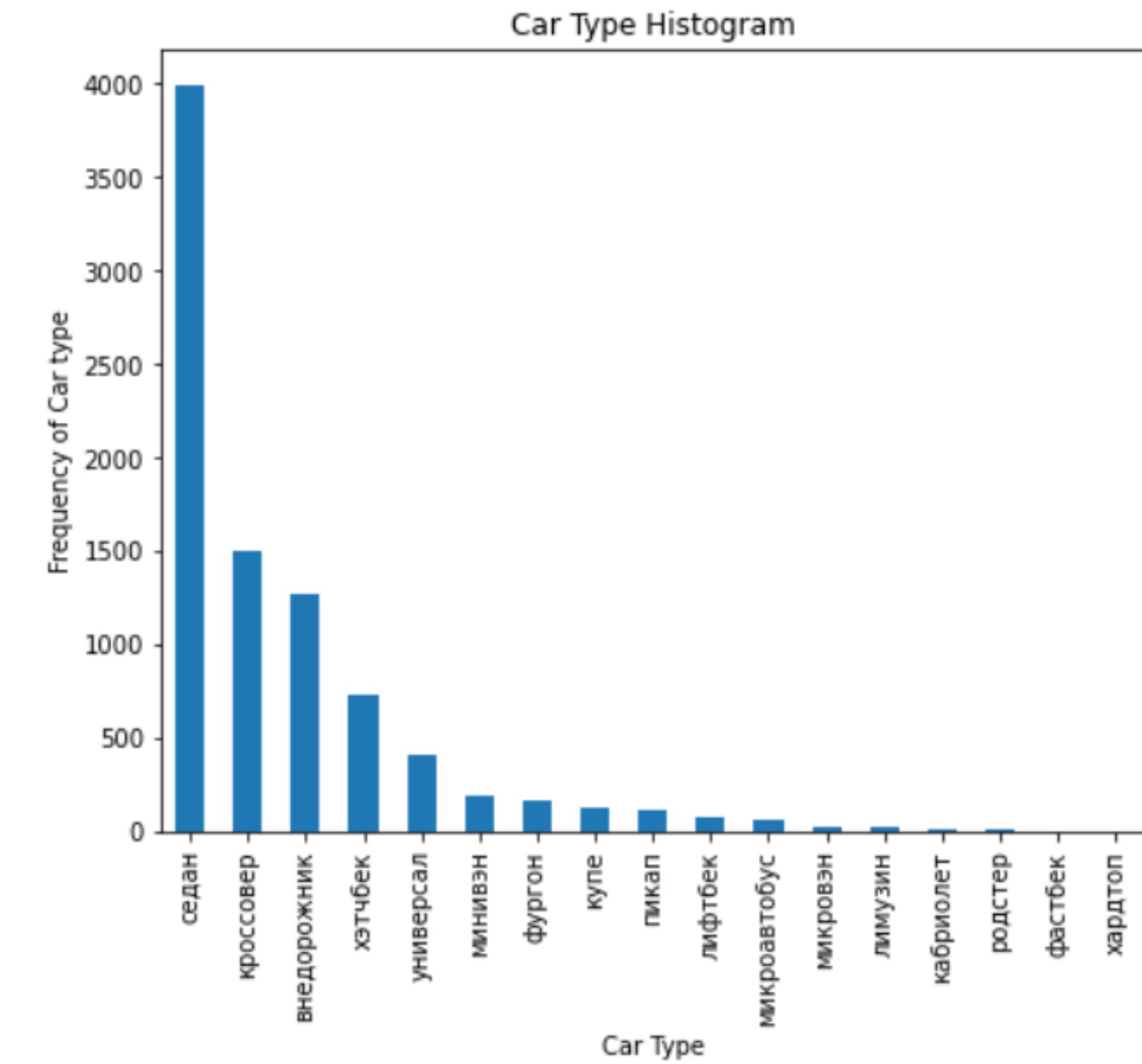
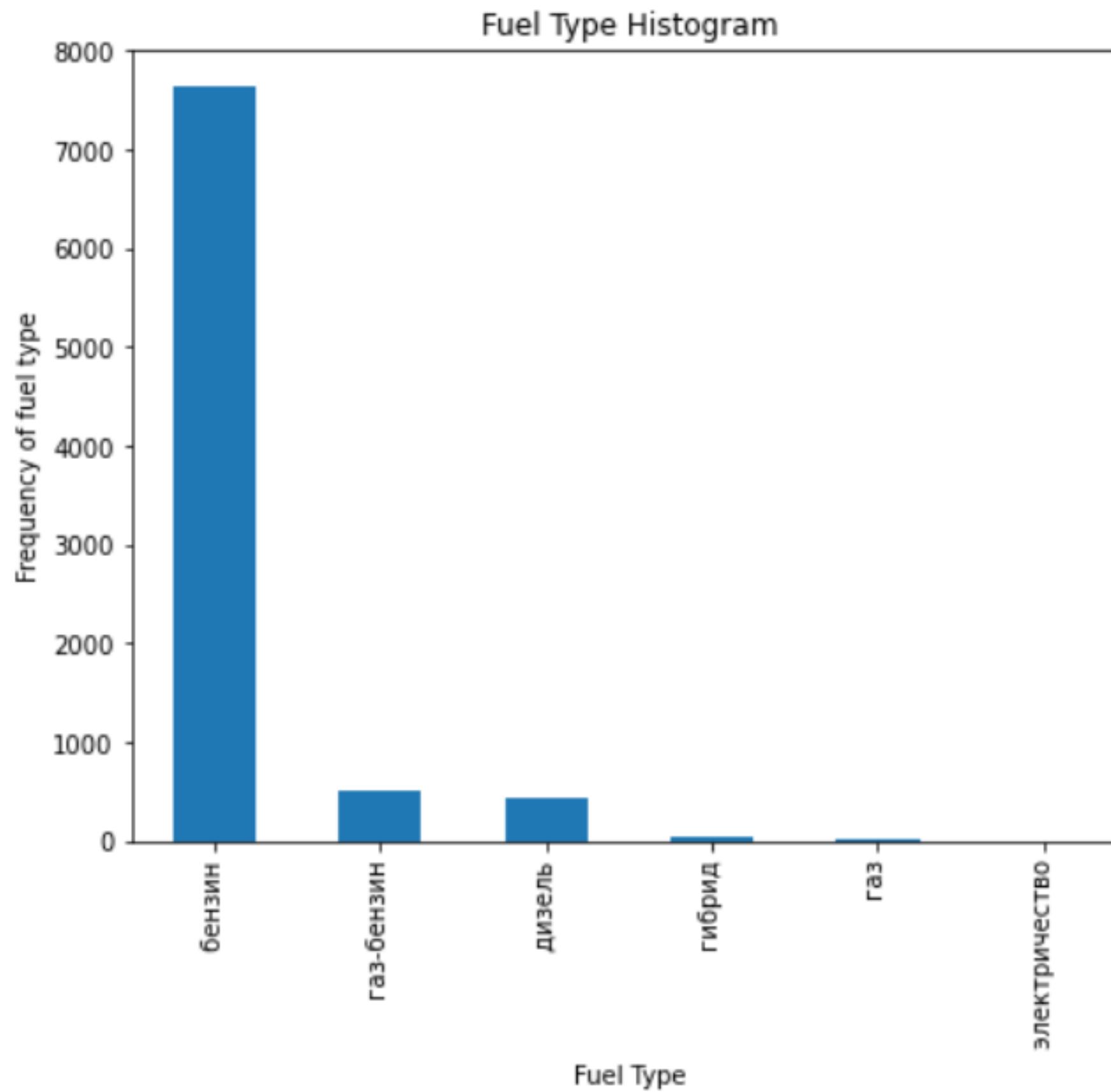
Car_id	City	Mark	Model	Price	Year	Body	EngineSize	FuelType	earboxTyp	Mileage	Color
1	Алматы	Toyota	Avensis	4200000	2003	седан	2	бензин	автомат	200000	серебристый
3	Алматы	Mitsubishi	Delica	12500000	2006	микроавтобус	3	бензин	автомат	180000	серебристый
4	Алматы	Infiniti	QX80	28050000	2018	внедорожник	5,6	бензин	автомат	22000	белый
6	Алматы	Nissan	Patrol	11500000	2012	внедорожник	5,6	бензин	автомат	105000	белый
8	Алматы	ВАЗ	(Lada)Largus	4800000	2019	минивэн	1,6	бензин	механика	90000	черный
10	Алматы	Volkswagen	Golf	1750000	1995	хэтчбек	2	бензин	механика	120000	синий
11	Алматы	Toyota	Corolla	3700000	2012	седан	1,4	бензин	механика	90000	серебристый
12	Алматы	Toyota	LandCruiser	9150000	2007	внедорожник	2,7	газ-бензин	автомат	196000	черный
13	Алматы	Mercedes-Benz	GLCCoupe	26500000	2019	кроссовер	2	бензин	автомат	4000	синий
17	Алматы	Toyota	Venza	8900000	2010	кроссовер	2,7	бензин	автомат	146000	серый
18	Алматы	ВАЗ	(Lada)2190	3200000	2015	седан	1,6	бензин	механика	18000	серебристый
22	Алматы	Toyota	LandCruiser	17500000	2008	внедорожник	4	бензин	автомат	118000	серый
26	Алматы	Hyundai	Accent	4000000	2012	седан	1,4	бензин	автомат	107000	черный
27	Алматы	Toyota	LandCruiser	23500000	2013	внедорожник	4,6	бензин	автомат	75000	белый
28	Алматы	Toyota	Camry	10100000	2015	седан	2	бензин	автомат	41000	белый
30	Алматы	Toyota	Camry	12200000	2017	седан	2,5	бензин	автомат	58000	белый
32	Алматы	Porsche	Panamera	65000000	2017	лифтбек	4	бензин	робот	12000	черный
34	Алматы	Infiniti	FX37	11000000	2012	кроссовер	3,7	бензин	автомат	145000	белый
36	Алматы	Toyota	Camry	12600000	2018	седан	2,5	бензин	автомат	51089	серебристый
39	Алматы	Toyota	LandCruiser	16800000	2011	внедорожник	4,7	бензин	типтроник	171000	белый
42	Алматы	Land Rover	RoverRange	13300000	2014	кроссовер	2	бензин	автомат	38700	белый
48	Алматы	Kia	Rio	3600000	2012	хэтчбек	1,6	бензин	механика	175000	серый
50	Алматы	Lexus	LX570	38900000	2016	внедорожник	5,7	бензин	типтроник	33000	белый
52	Алматы	Foton	Alpha	1000000	2005	фургон	2,2	дизель	механика	11111	белый
53	Алматы	Mazda	MPV	3800000	2004	минивэн	3	бензин	автомат	240000	серый
56	Алматы	Chevrolet	Spark	3950000	2018	хэтчбек	1	бензин	автомат	34000	белый
59	Алматы	Lexus	GS300	6000000	2007	седан	3	бензин	автомат	180000	черный
60	Алматы	Toyota	LandCruiser	7400000	2004	внедорожник	4,7	газ-бензин	автомат	346000	серебристый
61	Алматы	Hyundai	Accent	6000000	2019	седан	1,6	бензин	типтроник	42000	серебристый
64	Алматы	Toyota	LandCruiser	9200000	2007	внедорожник	2,7	бензин	автомат	235000	серебристый

Car Mark Frequency

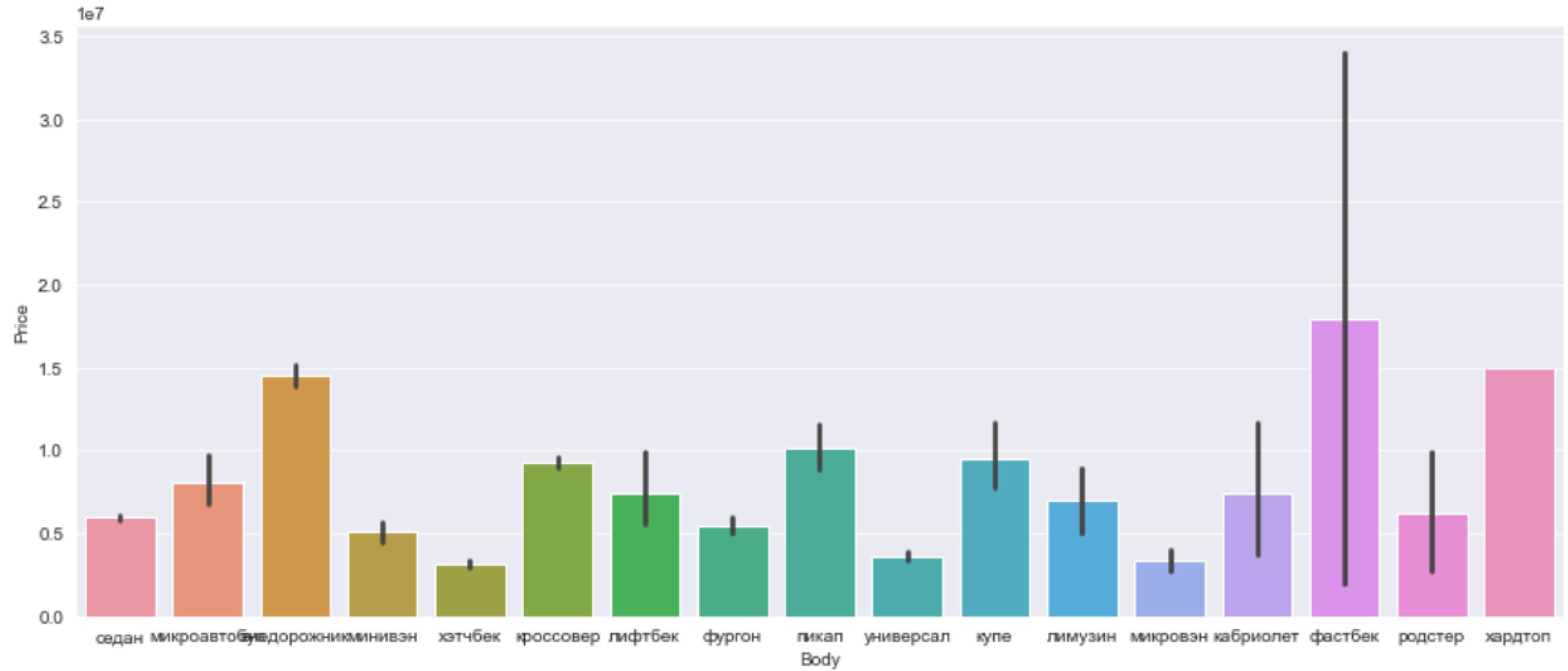
Cars Mark



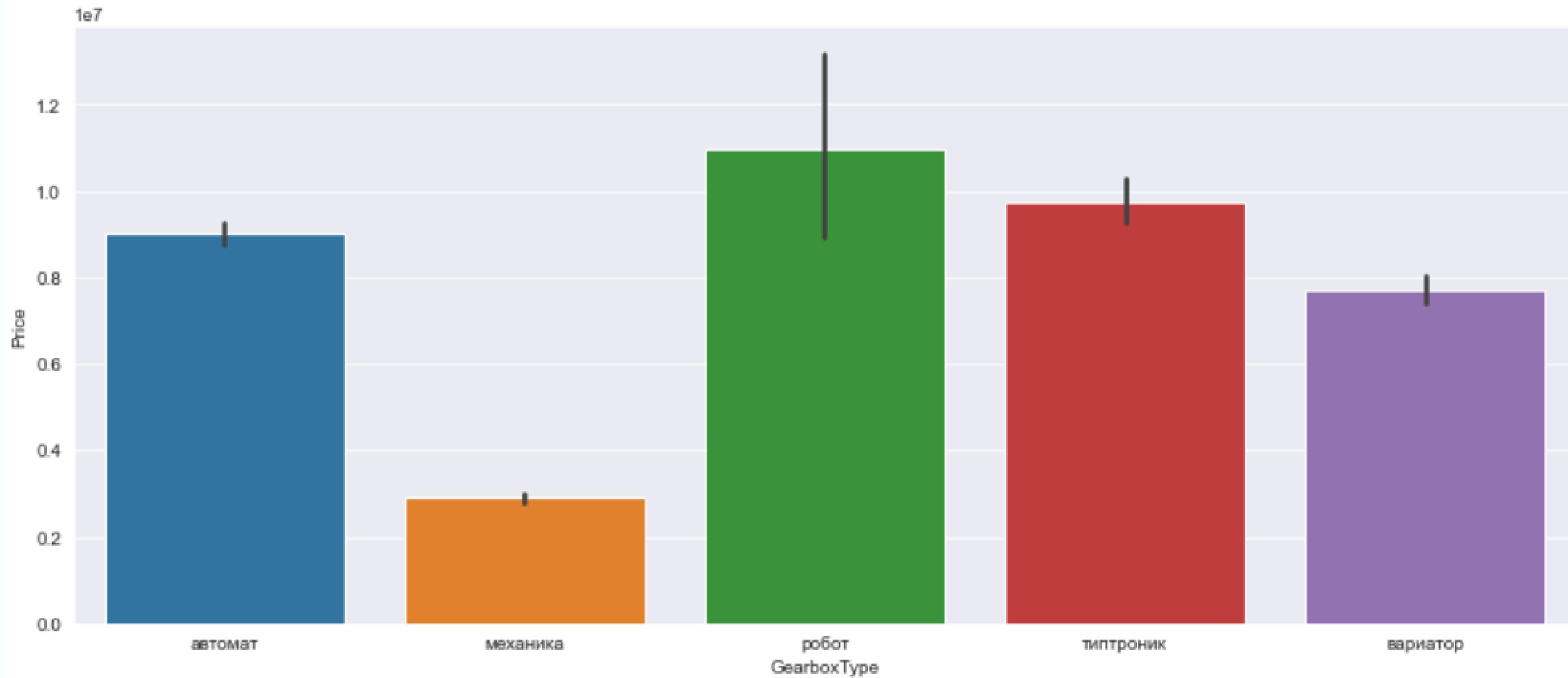
Fuel Type/Car Type Histogram



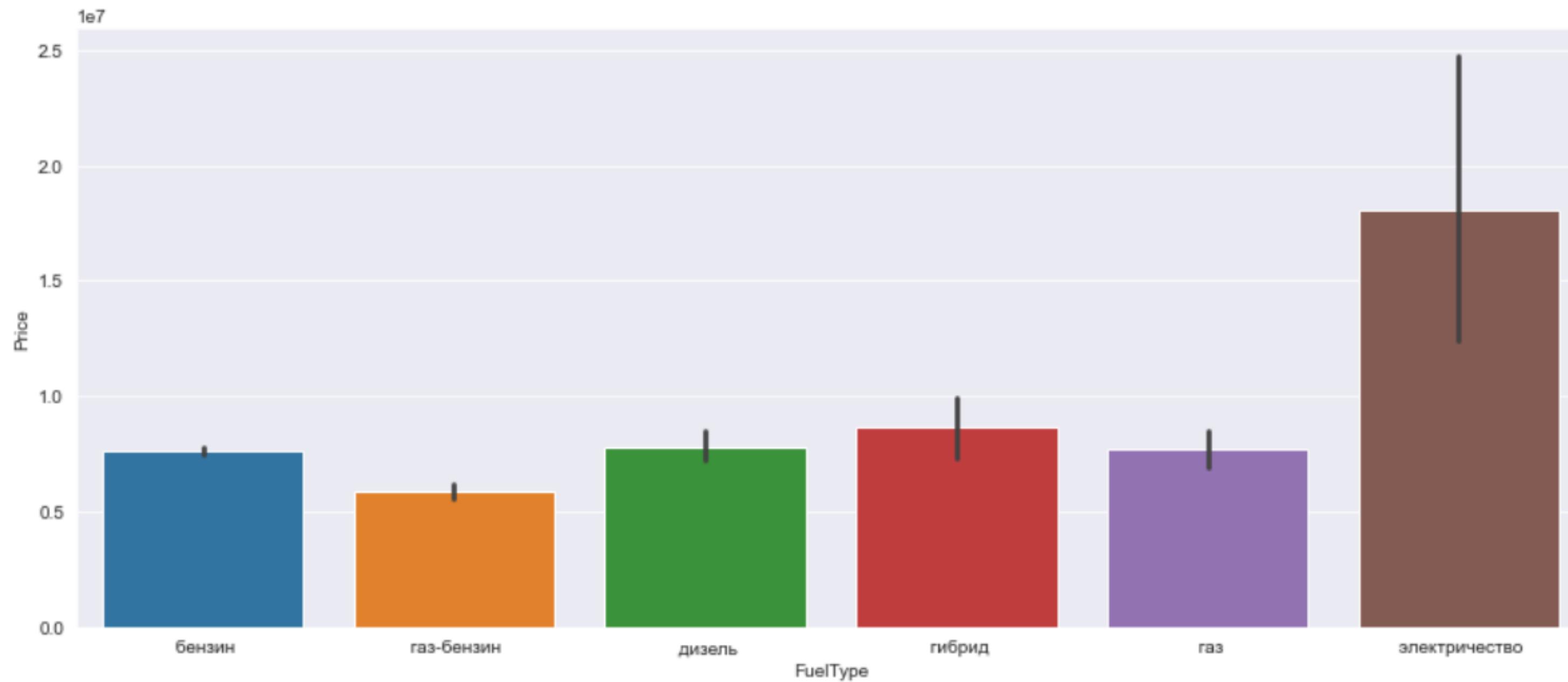
Car Body Average Prices



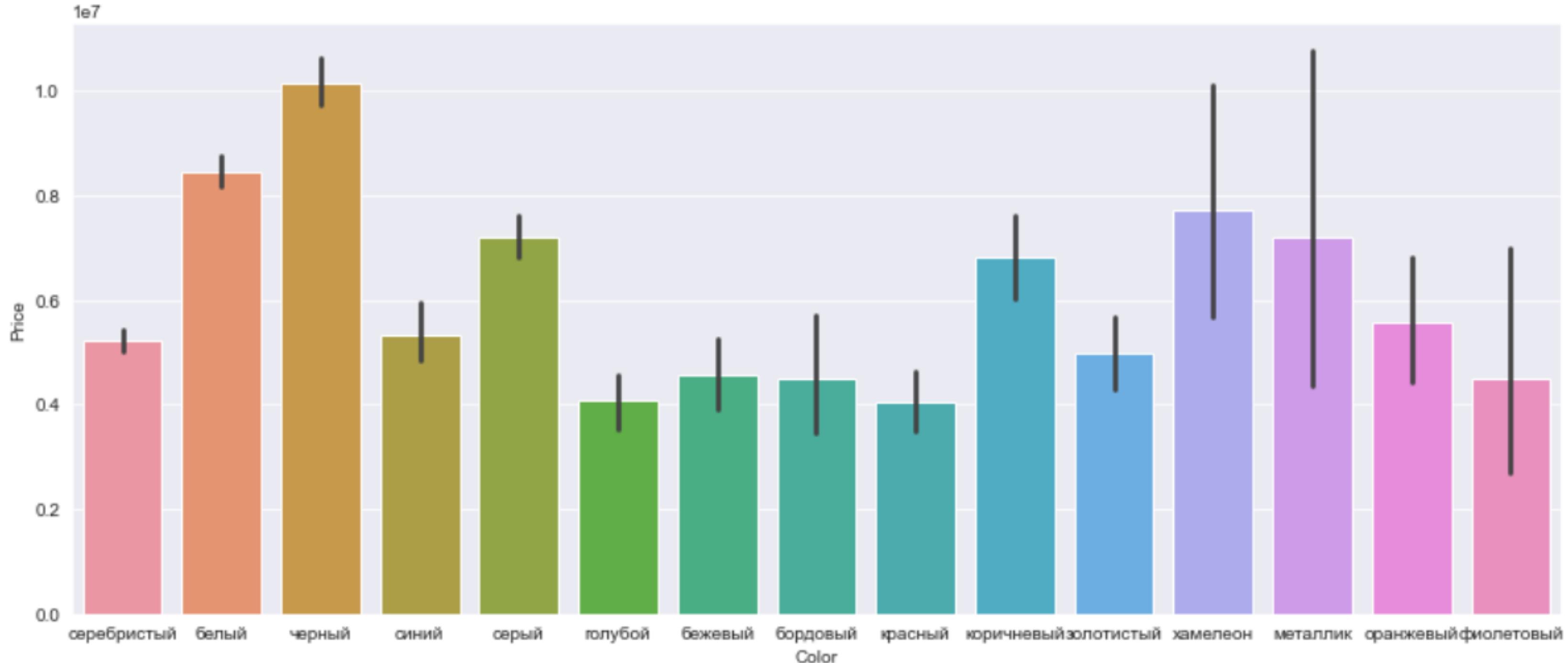
Car GearboxType Average Prices



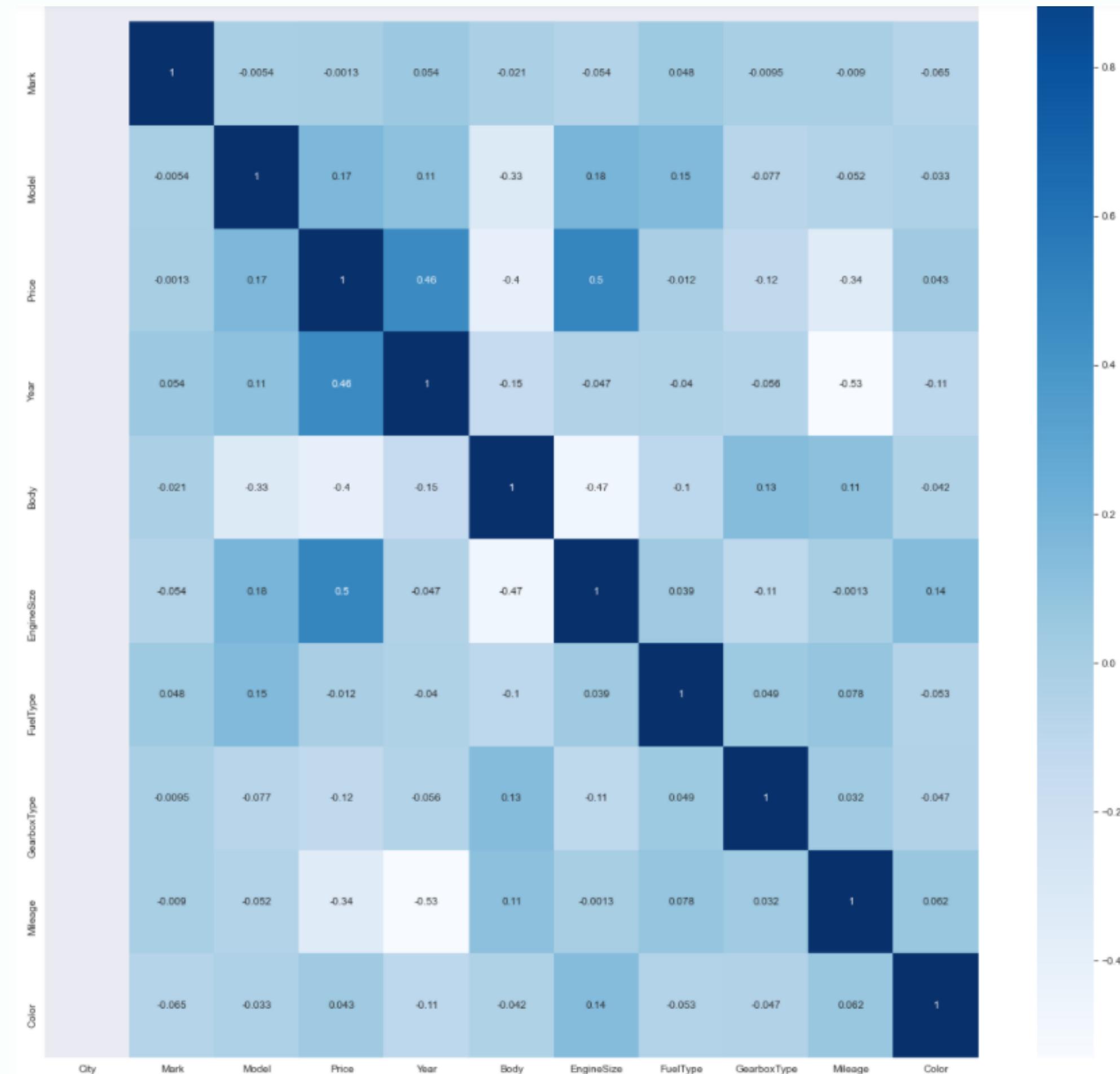
Car FuelType Average Prices



Car Color Average Prices



Heatmap





Algorithms



Random Forest
Linear Regression
Ridge Regression

Lasso
KNN
XGBoost

Random Forest

```
from sklearn.ensemble import RandomForestRegressor  
regressor = RandomForestRegressor(n_estimators = 10, random_state = 0)  
regressor.fit(X_train, y_train)
```

```
RandomForestRegressor(n_estimators=10, random_state=0)
```

```
y_pred=regressor.predict(X_test)
```

```
from sklearn.metrics import r2_score  
r2_score(y_test, y_pred)
```

```
0.8315636060029487
```

Accuracy: 83.16%

Random Forest

Predicted/ Actual Price

	Mark	Model	Year	Engine Size	Fuel Type	Body	Gearbox Type	Mileage	Color	actual_price	predictions
2527	Land Rover	Freelander	2014	2.0	бензин	кроссовер	автомат	59000	синий	9700000	13990000.0
6141	Toyota	Corolla	2011	1.6	бензин	седан	автомат	137500	белый	5400000	6225000.0
5268	Volkswagen	Golf	1997	1.8	бензин	хэтчбек	механика	240000	синий	1700000	1835000.0
7910	Subaru	Tribeca	2007	3.6	бензин	кроссовер	автомат	265983	белый	5250000	5839000.0
3182	Daewoo	Nexia	2013	1.5	бензин	седан	механика	190000	белый	1640000	1760000.0

Accuracy: 83.16%

Linear regression

```
from sklearn.linear_model import LinearRegression  
lr = LinearRegression()  
  
# Training Model  
lr.fit(x_train,y_train)  
  
# Model Summary  
y_pred_lr = lr.predict(x_test)  
  
r_squared = r2_score(y_test,y_pred_lr)  
rmse = np.sqrt(mean_squared_error(y_test,y_pred_lr))  
print("R_squared : ",r_squared)  
print("RMSE : ",rmse)
```

R_squared : 0.5567774964386283

RMSE : 5142146.302533223

Accuracy 55.68%

Linear regression

Predicted/ Actual Price

	Mark	Model	Year	EngineSize	FuelType	Body	GearboxType	Mileage	Color	actual_price	predictions
2527	Land Rover	Freelander	2014	2.0	бензин	кроссовер	автомат	59000	синий	9700000	9.728731e+06
6141	Toyota	Corolla	2011	1.6	бензин	седан	автомат	137500	белый	5400000	5.078285e+06
5268	Volkswagen	Golf	1997	1.8	бензин	хэтчбек	механика	240000	синий	1700000	-1.114986e+06
7910	Subaru	Tribeca	2007	3.6	бензин	кроссовер	автомат	265983	белый	5250000	1.045664e+07
3182	Daewoo	Nexia	2013	1.5	бензин	седан	механика	190000	белый	1640000	4.823511e+06

Accuracy 55.68%

Ridge regression

```
from sklearn.linear_model import Ridge  
  
rdg=Ridge(alpha=1.0)  
rdg.fit(x_train,y_train)  
  
y_pred_rdg=rdg.predict(x_test)  
  
r_squared = r2_score(y_test,y_pred_rdg)  
rmse = np.sqrt(mean_squared_error(y_test,y_pred_rdg))  
print("R_squared : ",r_squared)  
print("RMSE : ",rmse)
```

R_squared : 0.5567716056233141
RMSE : 5142180.474234

Accuracy 55%

Ridge regression

Predicted/ Actual Price

	Mark	Model	Year	EngineSize	FuelType	Body	GearboxType	Mileage	Color	actual_price	predictions
2527	Land Rover	Freelander	2014	2.0	бензин	кроссовер	автомат	59000	синий	9700000	9.729345e+06
6141	Toyota	Corolla	2011	1.6	бензин	седан	автомат	137500	белый	5400000	5.078495e+06
5268	Volkswagen	Golf	1997	1.8	бензин	хэтчбек	механика	240000	синий	1700000	-1.114896e+06
7910	Subaru	Tribeca	2007	3.6	бензин	кроссовер	автомат	265983	белый	5250000	1.045640e+07
3182	Daewoo	Nexia	2013	1.5	бензин	седан	механика	190000	белый	1640000	4.823826e+06

Accuracy 55%

Lasso

```
In [39]: from sklearn.linear_model import Lasso  
from sklearn.metrics import mean_squared_error  
  
lso = Lasso(alpha=1.0)  
lso.fit(X_train, y_train)  
y_pred_lso = lso.predict(X_test)
```

```
In [40]: from sklearn.metrics import r2_score  
r2_score(y_test, y_pred_lso)
```

Out[40]: 0.5567774786106191

Accuracy 55%

Lasso

Predicted/ Actual Price

	Mark	Model	Year	EngineSize	FuelType	Body	GearboxType	Mileage	Color	actual_price	predictions
2527	Land Rover	Freelander	2014	2.0	бензин	кроссовер	автомат	59000	синий	9700000	9.728732e+06
6141	Toyota	Corolla	2011	1.6	бензин	седан	автомат	137500	белый	5400000	5.078285e+06
5268	Volkswagen	Golf	1997	1.8	бензин	хэтчбек	механика	240000	синий	1700000	-1.114986e+06
7910	Subaru	Tribeca	2007	3.6	бензин	кроссовер	автомат	265983	белый	5250000	1.045664e+07
3182	Daewoo	Nexia	2013	1.5	бензин	седан	механика	190000	белый	1640000	4.823512e+06

Accuracy 55%

KNN

```
In [32]: from sklearn.neighbors import KNeighborsRegressor  
  
knn=KNeighborsRegressor(n_neighbors=7)  
knn.fit(X_train, y_train)  
y_pred_knn=knn.predict(X_test)
```

```
In [33]: from sklearn.metrics import r2_score  
r2_score(y_test, y_pred_knn)
```

Out[33]: 0.24200554077381542

Accuracy 24.2%

KNN

Predicted/ Actual Price

	Mark	Model	Year	EngineSize	FuelType	Body	GearboxType	Mileage	Color	actual_price	predictions
2527	Land Rover\ltFreelander		2014	2.0	бензин	кроссовер	автомат	59000	синий	9700000	1.645000e+07
6141	Toyota	Corolla	2011	1.6	бензин	седан	автомат	137500	белый	5400000	4.870000e+06
5268	Volkswagen	Golf	1997	1.8	бензин	хэтчбек	механика	240000	синий	1700000	5.735714e+06
7910	Subaru	Tribeca	2007	3.6	бензин	кроссовер	автомат	265983	белый	5250000	3.398571e+06
3182	Daewoo	Nexia	2013	1.5	бензин	седан	механика	190000	белый	1640000	5.321429e+06

Accuracy 24.2%

XGBoost

```
In [24]: from xgboost import XGBRegressor  
  
xgb = XGBRegressor(n_estimators=1000, learning_rate=0.1)  
xgb.fit(X_train, y_train, early_stopping_rounds=5, eval_set=[(X_test, y_test)], verbose=False)  
y_pred_xgb = xgb.predict(X_test)
```

```
In [25]: from sklearn.metrics import r2_score  
r2_score(y_test, y_pred_xgb)
```

Out[25]: 0.9031754676787723

Accuracy 90.3%

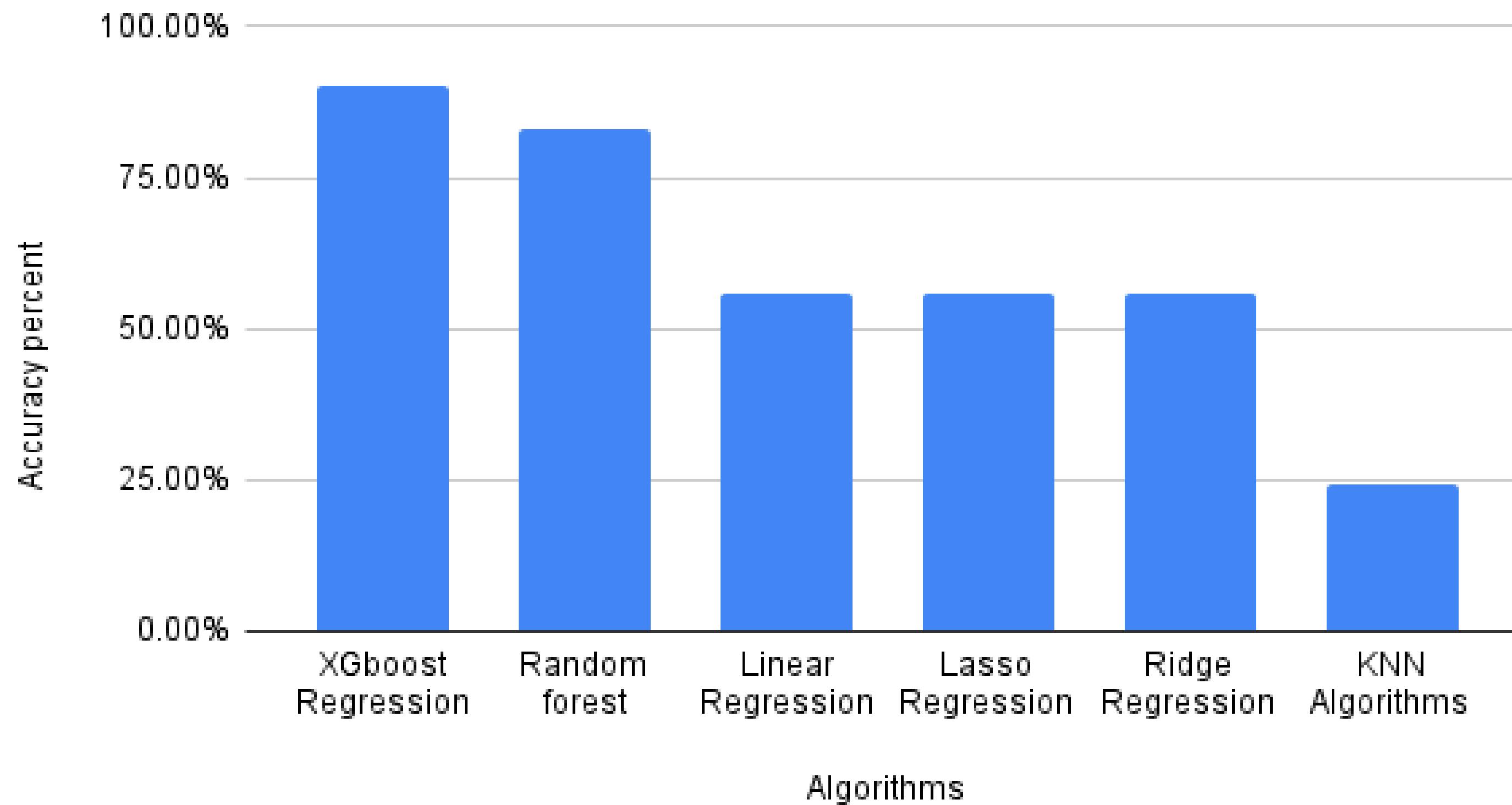
XGBoost

Predicted/ Actual Price

	Mark	Model	Year	Engine Size	Fuel Type	Body	Gearbox Type	Mileage	Color	actual_price	predictions
2527	Land Rover	Freelander	2014	2.0	бензин	кроссовер	автомат	59000	синий	9700000	1.464440e+07
6141	Toyota	Corolla	2011	1.6	бензин	седан	автомат	137500	белый	5400000	5.460000e+06
5268	Volkswagen	Golf	1997	1.8	бензин	хэтчбек	механика	240000	синий	1700000	1.817381e+06
7910	Subaru	Tribeca	2007	3.6	бензин	кроссовер	автомат	265983	белый	5250000	5.880989e+06
3182	Daewoo	Nexia	2013	1.5	бензин	седан	механика	190000	белый	1640000	1.612158e+06

Accuracy 90.3%

Accuracy percent относительно параметра "Algorithms"



Analyzing Abnormal Data

Normal Data RF

```
In [4]: value = 0.13
```

```
In [5]: Random_forest_data_normal = Random_forest_data.query("predictions < actual_price + (@value*actual_price) & \
                                                               predictions > actual_price - (@value*actual_price)")
Random_forest_data_normal
```

Out[5]:

	car_id	Mark	Model	Year	EngineSize	FuelType	Body	GearboxType	Mileage	Color	actual_price	predictions
2	5268	Volkswagen	Golf	1997	1.8	бензин	хэтчбек	механика	240000	синий	1700000	1835000.0
3	7910	Subaru	Tribeca	2007	3.6	бензин	кроссовер	автомат	265983	белый	5250000	5839000.0
4	3182	Daewoo	Nexia	2013	1.5	бензин	седан	механика	190000	белый	1640000	1760000.0
7	3236	Nissan	Qashqai	2007	2.0	бензин	кроссовер	вариатор	240550	коричневый	4620000	4565000.0
8	5989	Volkswagen	Passat	1992	1.8	бензин	универсал	механика	290000	бордовый	1550000	1395400.5
...
1726	3796	Toyota	LandCruiser	2012	4.6	бензин	внедорожник	автомат	145000	черный	20800000	21515000.0
1727	3847	BMW	318	2002	1.9	бензин	купе	механика	350350	серебристый	2800000	2699000.0
1728	8630	ГАЗ	ГАЗельNEXT	2018	2.7	бензин	фургон	механика	120000	белый	8499999	8044700.0
1729	5003	Toyota	LandCruiser	2008	4.7	газ-бензин	внедорожник	автомат	263000	белый	14600000	13250000.0
1730	4902	Nissan	Patrol	2013	5.6	бензин	внедорожник	автомат	222128	черный	12820000	13095000.0

948 rows × 12 columns

Abnormal Data RF

```
In [6]: Random_forest_data_notnormal = Random_forest_data.query("predictions > actual_price + (@value*actual_price) | \
                                                               predictions < actual_price - (@value*actual_price)")
Random_forest_data_notnormal
```

Out[6]:

	car_id	Mark	Model	Year	EngineSize	FuelType	Body	GearboxType	Mileage	Color	actual_price	predictions
0	2527	Land Rover	Freelander	2014	2.0	бензин	кроссовер	автомат	59000	синий	9700000	13990000.0
1	6141	Toyota	Corolla	2011	1.6	бензин	седан	автомат	137500	белый	5400000	6225000.0
5	3285	Porsche	Cayenne	2008	4.8	бензин	кроссовер	типтроник	109000	голубой	7600000	9310000.0
6	8576	Toyota	HiAce	2012	2.7	бензин	микроавтобус	механика	226000	серебристый	8900000	10178998.0
9	1696	Hyundai	Coupe	2004	2.0	бензин	купе	автомат	155000	красный	2900000	3335000.0
...
1719	8181	Toyota	Land Cruiser	2015	4.6	бензин	внедорожник	автомат	49700	белый	32500000	24590000.0
1720	3133	ВАЗ (Lada)	2121 Нива	2004	1.7	бензин	внедорожник	механика	170000	белый	1700000	1069555.3
1723	8545	Mitsubishi	Galant	2005	2.4	бензин	седан	автомат	429655	серый	2500000	4304000.0
1724	526	Volkswagen	Transporter	2007	2.5	дизель	минивэн	механика	173000	белый	8000000	5549000.0
1731	313	Kia	Cerato Koup	2012	2.0	бензин	купе	типтроник	100000	серебристый	4900000	5550000.0

782 rows × 12 columns

Algorithms	Normal	Abnormal
XGboost Regression	944	788
Random forest	948	782
Linear Regression	387	1345
Lasso Regression	387	1345
Ridge Regression	387	1345
KNN Algorithms	256	1476

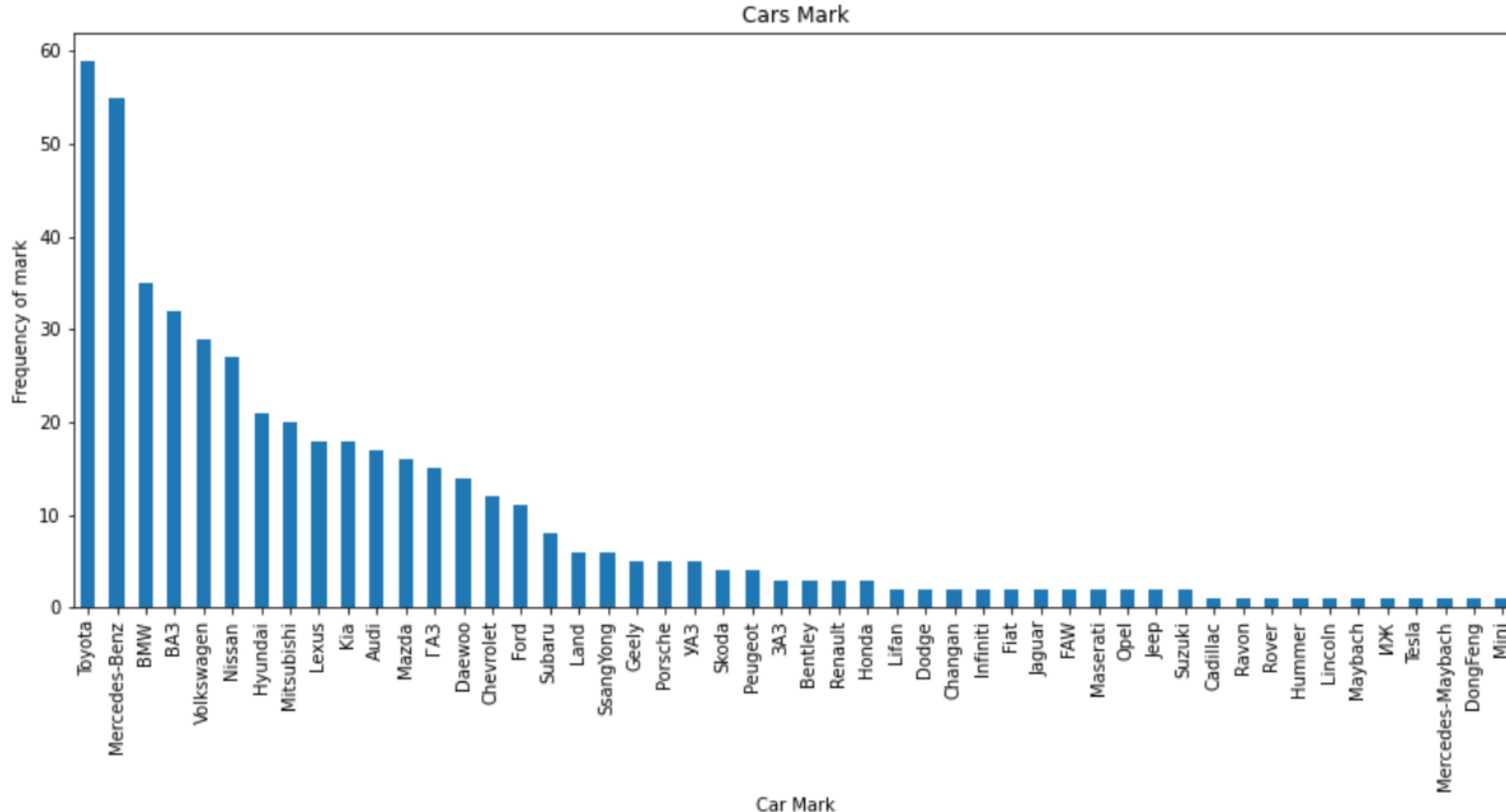
Intersections of abnormal values between all models

	car_id	Mark	Model	Year	EngineSize	FuelType	Body	GearboxType	Mileage	Color	actual_price	predictions
5	3285	Porsche	Cayenne	2008	4.8	бензин	кроссовер	типтроник	109000	голубой	7600000	8628735.00
12	4285	Mercedes-Benz	S\lt450	2006	4.5	бензин	седан	автомат	293196	серебристый	7000000	5823991.50
16	7084	BMW	X6	2012	4.4	бензин	кроссовер	типтроник	76000	черный	10500000	16097132.00
18	5071	Hyundai	Grandeur	2016	3.0	бензин	седан	автомат	82000	серебристый	10000000	17663204.00
21	4718	ВАЗ (Lada)\t2121\тНива	2013		1.7	бензин	внедорожник	механика	45896	белый	2450000	3134271.25
...
1708	1974	Audi	80	1990	2.0	бензин	седан	механика	380000	красный	1380000	1041968.50
1717	4230	Kia	Cee'd	2013	1.6	бензин	хэтчбек	типтроник	253000	коричневый	5144000	4256721.00
1719	8181	Toyota	Land\ltCruiser	2015	4.6	бензин	внедорожник	автомат	49700	белый	32500000	26162330.00
1724	526	Volkswagen	Transporter	2007	2.5	дизель	минивэн	механика	173000	белый	8000000	4772302.00
1731	313	Kia	Cerato\ltKoup	2012	2.0	бензин	купе	типтроник	100000	серебристый	4900000	6880113.50

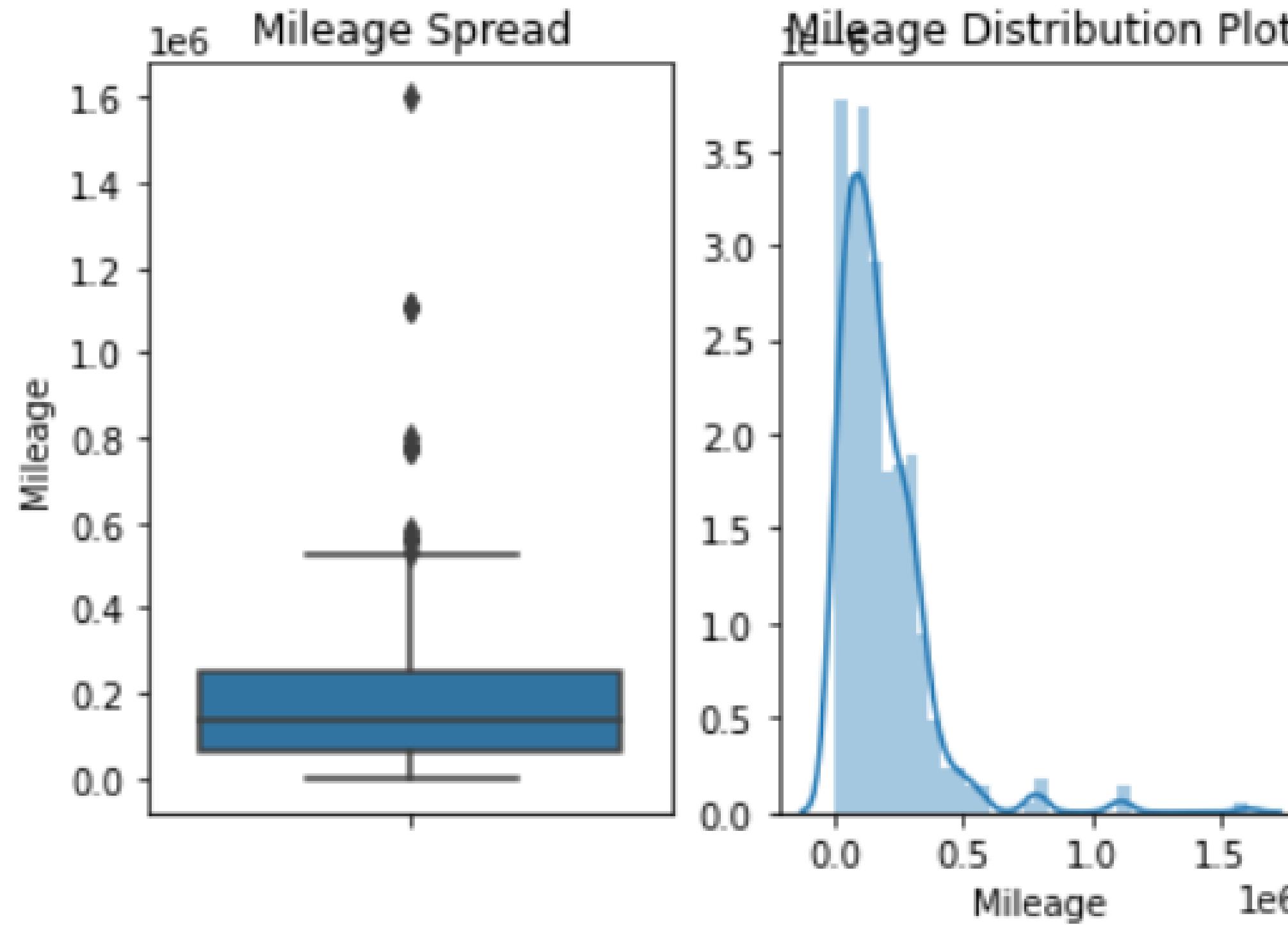
420 rows × 12 columns

Intersection: 420

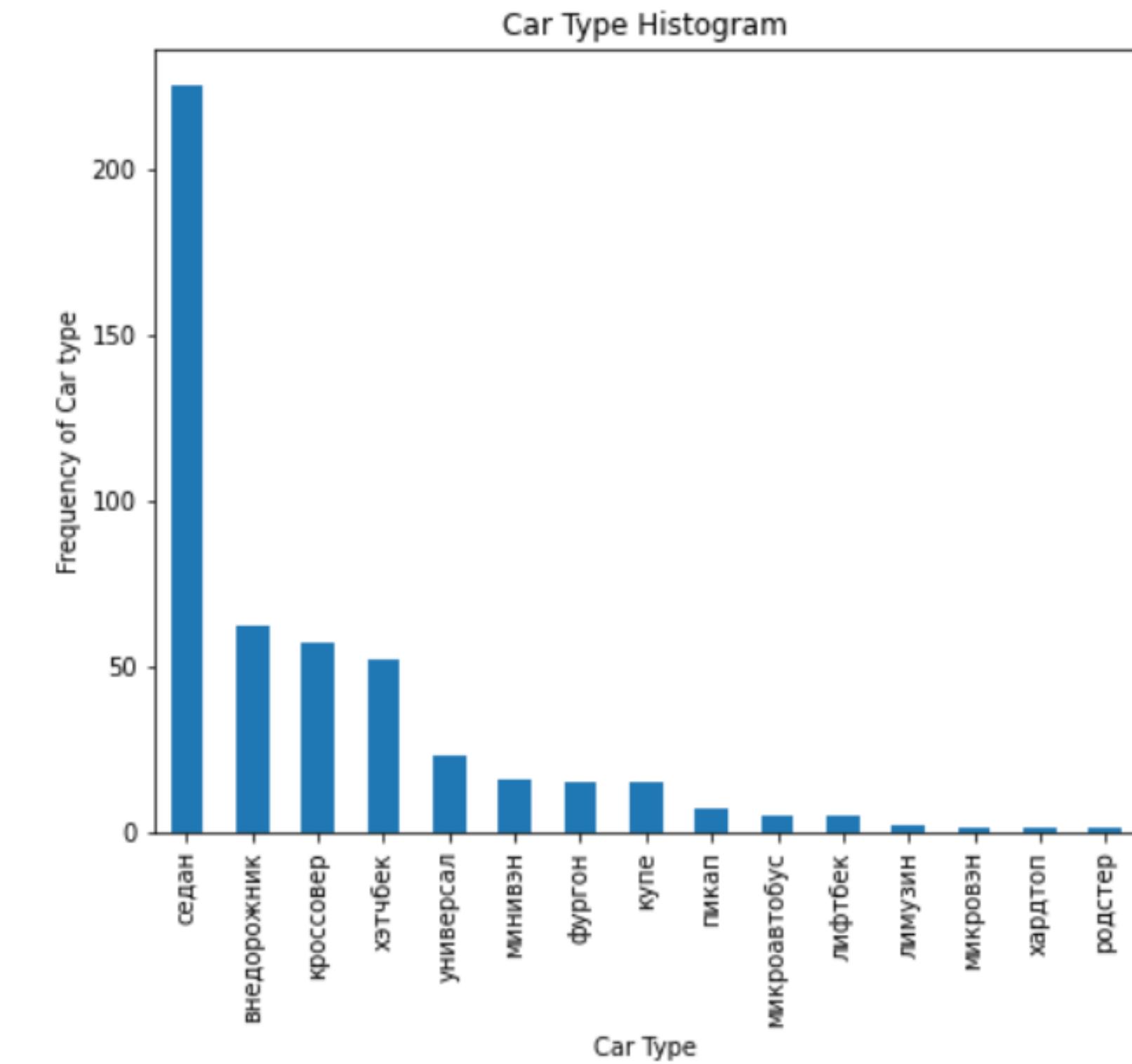
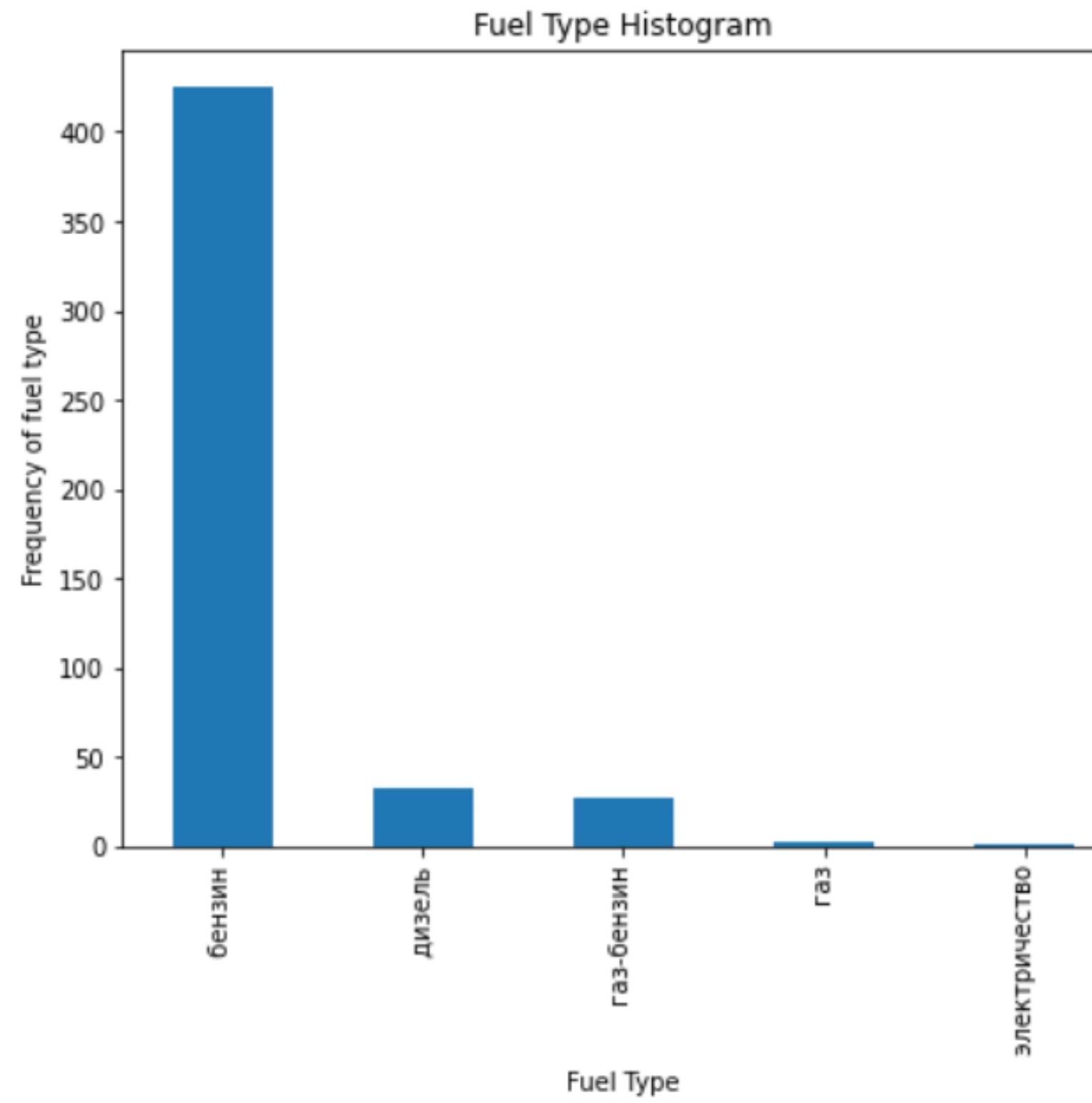
Car frequency in abnormal dataframe



Car Mileage Spread in abnormal dataframe



Car Fuel Type/ Car Type Histogram Spread in **abnormal** dataframe



How to improve our model accuracy?

- 1) Use more category data**
- 2) Add more data for prediction model**

Thank you for your attention!



```
(all_cars/cars).sort_values(ascending=False).head(25).to_frame()
```

Mark	
Honda	17.000000
Toyota	8.590909
Lexus	8.000000
Renault	7.000000
Hyundai	6.529412
Kia	6.285714
Skoda	6.000000
Land	5.800000
Subaru	5.750000
Infiniti	5.500000
Nissan	4.333333
Cadillac	4.000000
Ravon	4.000000
Mercedes-Benz	3.702128
Mitsubishi	3.611111
Suzuki	3.500000
Opel	3.500000
Volkswagen	3.296296
Porsche	3.250000
BA3	3.233333
BMW	3.032258
Peugeot	2.750000
Audi	2.705882
Chevrolet	2.583333
Jeep	2.500000