1. Wynik Scrappingu:

Program scrapuje dane z serwisu Otomoto. Dane te są pobierane tylko dla aut osobowych.

Zmienne jakie program pobiera to:  
- Brand – marka samochodu. Jeśli marka nie zostaje dopasowana do bazy nazw marek, dostaje kategorie „Inny”,

- Model – model samochodu podany przez użytkownika w ogłoszeniu (bez nazwy marki). Jeżeli marka samochodu nie została dopasowana do klucza to nazwa modelu pozostaje wraz z nazwą marki,

- Price – cena samochodu wyrażona w złotówkach bez części groszowej,

- Engine – pojemność silnika,

- KM – Konie mechaniczne; moc silnika,

- Mileage – przebieg samochodu wyrażony w km,

- Fuel – zastosowany sposób zasilania samochodu,

- Gearbox – Zastosowana skrzynia w samochodzie. Rozróżnialna jest tylko skrzynia manualna lub automatyczna,

- Year – rok produkcji

- Place – lokalizacja samochodu

- Voivode – województwo w którym jest ogłoszony samochód

Wynik jest zapisywany w pliku json i pliku csv, który jest używany do analizy danych.

1. Założenia jakie program ma osiągnąć w analizie danych.

- podstawowe obliczenia na cenach czy województwach.

- korelacja pomiędzy ceną a zmiennymi technicznymi pojazdu.

- popularność danego rozwiązania technicznego

- wykres rozkładu cen

- tabela krzyżowa

- regresja liniowa ceny

3. Predykcja programu ma na celu ukazanie tendencji i przewidzenia cen pojazdów na podstawie ich danych technicznych oraz ceny.

4. Przed każdą analizą czy predykcją dane należało oczyścić i ustandaryzować:

- województwa Polski zostały odsiane od innych pozostałych miejsc w których ogłoszenie zostało dodane. Inne miejsca, które nie są województwem zostały podpisane jako „Inny kraj”.

- wszystkie dane liczbowe zostały przeparsowane na Integer. Dodatkowo jeżeli dany rekord nie posiadał wartości liczbowej, to zamiast Nan zostało wpisane 0.

- dane nie posiadające wartości zostały usunięte

- jeżeli do wykresu potrzebne były odpowiednie dla wykresu limitery lub inne działania na danych w celu lepszego przedstawienia wykresu, dane zostały odpowiednie zmodyfikowane.

- marka została wyciągnięta ze stringa podanego w modelu poprzez zastosowanie dopasowania z filtrem nazwy z pliku yaml.

- w przypadku predykcji dane kategoryczne zostały zamienione przy pomocy funkcji get\_dummies zamienione na osobne kolumny.

5. Wyniki podstawowych obliczeń danych.

a) Średnia cena samochodu jest równa 78247 zł. Możliwe że średnia została mocno wyciągnięta do góry przez drogie nowe samochody z marek premium.

b) Maksymalna cena: Mercedes-Benz Klasa G AMG 63 za prawie 5 milionów jest przykładem dlaczego średnia jest aż tak wysoka.

c) Minimalna cena: 500 dla Volkswagena Scirocco 1.4 TSI. Nie było możliwe od sortować samochodów nie działających, po wypadkach, itp. Dlatego nie wiadomo czy ten samochód za 500 zł to rzeczywiście najtańszy dostępny jeżdżący samochód.

d) Korelacja pomiędzy ceną a pojemnością silnika, rokiem produkcji, mocą silnika, przebiegiem. Wyniki to odpowiednio (wyniki w analizie są zaokrąglane do dwóch miejsc po przecinku):

- Poj. silnika – -0,13 – raczej nikły wpływ

- Rok produkcji – 0,51 – znaczny wpływ. Ma to sens, ponieważ zwykły nowy samochód z 2025 będzie droższy niż samochód z 2010.

- Moc silnika – 0,64 – duży wpływ. Im silnik mocniejszy tym cena jest większa co ma przełożenie na codzienne doświadczenie.

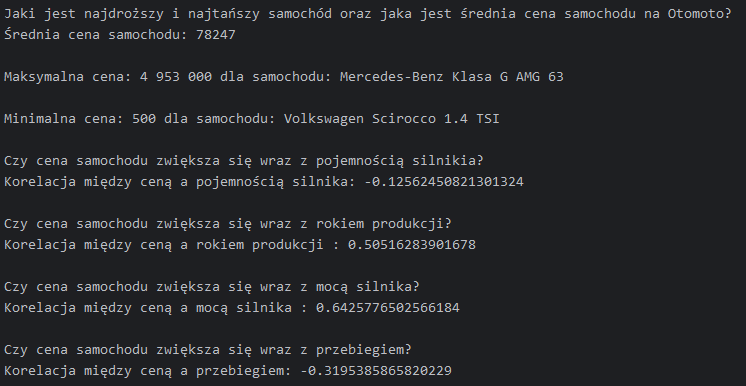
- Przebieg - -0,32 – Im mniejszy przebieg tym większa cena. Wynik jest mały ale ma wpływ na cenę co również ma rację bytu. Nowy samochód jest droższy niż używany, a posiada bardzo mały przebieg. Mimo wszystko, wydaje się, że przebieg powinien mieć duże przełożenie na cenę a w rzeczywistości tak nie jest.

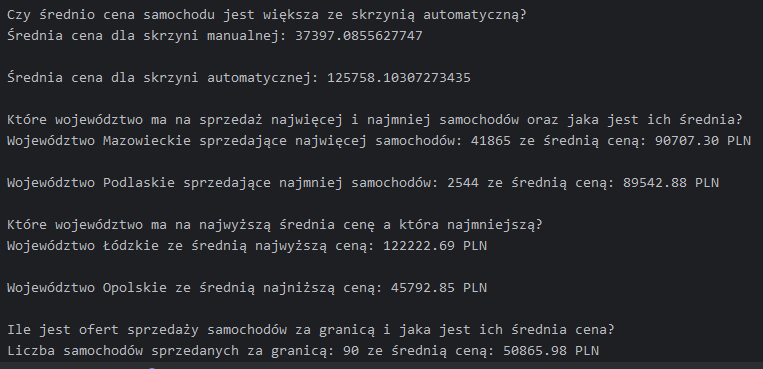
e) Które samochody są średnio droższe ze skrzynią biegów manualną czy automatyczną?

37397 < 125758  
To bardzo duża różnica, więc z całą pewnością można pwoeidzieć, że samochody ze skrzynią biegów automatyczną jest znacznie droższy niż manualny odpowiednik.

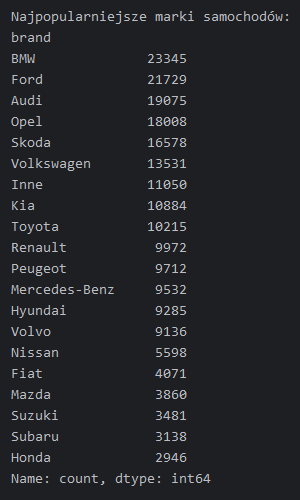
f) Województwo Mazowieckie sprzedaje najwięcej aut bo aż 42000 w przybliżeniu, natomiast województwo Podlaskie sprzedaje najmniej bo tylko 2500. To aż 95% mniej niż w Mazowszu.

g) na polskim serwisie ogłoszeniowym Otomoto, jest 90 samochodów sprzedawanych za granicą z różnych krajów.





6. Popularność samochodów



Pierwsze 6 pozycji dominuje przez niemieckie koncerny samochodowe. Pokazuje to, że najczęściej sprzedawane są niemieckie maszyny.

Sekcja Inne jest na 7 pozycji, to wskazuje że filtr marek samochodów jest zbyt ogólny i pomimo że marek zostało wypisanych około 80. To ponad 10000 aut nie zostało przypisane do żadnej marki.

Ciekawe też jest obserwacja że numer 19 to Subaru. Wydaje się że ta japońska marka nie jest zbyt popularna w Polsce a jednak 3000 aut jest na sprzedaż. Pytanie czy to z powodu, że się nie sprzedają czy jednak Subaru wcale nie jest tak niepopularną marką.

Najpopularniejsze typy zasilania aut to benzyna. Aut elektrycznych jest 2000. Przy trochę ponad 200000 samochodów spalinowych to przepaść i jednak nowe elektryczne auta nie są aż tak popularne w sprzedaży.

Ciekawostką są ostatnie dwie pozycje, czyli samochód na Etanol i wodór, czyli auta które są bardzo niszowe to 11 samochodów jest na sprzedaż na Otomoto.