

Анализ данных о ценах на автомобили в Австралии

Екимова Яна Сергеевна

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Данная научная статья посвящена проведению анализа данных о ценах на автомобили в Австралии с использованием Google Colaboratory. Для достижения поставленной цели использовался набор данных, содержащий информацию о ценах, марке, модели, годе выпуска, технических характеристиках и других атрибутах автомобилей. Используя Python и библиотеки анализа данных, такие как Pandas, Plotly.express, Plotly.graph_objects был проведен анализ данных.

Ключевые слова: Google Colaboratory, Pandas, анализ данных.

Analysis of car price data in Australia

Ekimova Yana Sergeevna

Sholom Aleichem Priamurskiy State University

Student

Abstract

This scientific article is devoted to the analysis of data on car prices in Australia using Google Colaboratory. To achieve this goal, a data set containing information about prices, make, model, year of manufacture, technical characteristics and other attributes of cars was used. Using Python and data analysis libraries such as Pandas, Plotli.express, Plot.graph_objects data analysis was performed.

Key words: Google Colaboratory, Pandas, data analysis.

1. Введение

1.1. Актуальность

Google Colab остается актуальным и популярным выбором для проведения исследований, разработки моделей машинного обучения и выполнения вычислений в облаке. В настоящее время цены на автомобили являются важной и актуальной темой, поскольку автомобильная индустрия продолжает развиваться, и потребители все более осознанно подходят к покупке автомобиля. Анализ данных о ценах на автомобили позволяет выявить закономерности и факторы, влияющие на ценообразование, что полезно для покупателей и продавцов.

1.2 Обзор исследований

А.Р.Сухинина, Е.В.Королева, Е.А.Ефимов, Н.А. Благой в исследовании продемонстрировали зависимость доступности жилья от объемов ввода жилья. Темпов прироста населения, суммарного коэффициента рождаемости, коэффициента брачности, уровня занятости, удельного веса числа семей, получивших жилые помещения и улучшивших жилищные условия, количества предоставленных ипотечных жилищных кредитов, средневзвешенного срока и средневзвешенной процентной ставки по ипотечным жилищным кредитам [1]. Н.Е. Косых предложил подход к подбору наиболее производительного набора параметров для объекта классификатора текста. Для вычислений использовал облачный сервис Google Colaboratory, выполняющий код на языке Python внутри браузера, используя виртуальные аппаратные ресурсы [2]. Т.М. Татарникова, Е.Д.Пойманова, П.Ю.Богданов, Е.В.Краева, С.А. Веревкин в статье рассмотрели способы и методы изучения и построения нейронных сетей. Показано, что изучение принципов функционирования нейронных сетей, их применение для решения тех или иных задач возможны только через практику. Проведен анализ различных программных сред, которые могут быть использованы на лабораторных и практических занятиях по изучению и применению нейронных сетей. Выделен современный облачный сервис Google Colaboratory, рекомендуемый для обучения основам нейронных сетей благодаря наличию в нем предустановки библиотеки Tensorflow и библиотеки для работы на языке Python, бесплатного доступа к графическим процессорам, возможности написания и выполнения программного кода в браузере, а также отсутствию необходимости специальной настройки сервиса [3]. Е.А.Григорьев, Н.С.Климов в статье рассматривали разведочный анализ данных. Описали инструменты реализации анализа, библиотеки Python. Представлен пример выполненный на данных обнаружению присутствия людей в помещении [4].

1.3 Цель исследования

Цель исследования – выполнить анализ данных с помощью Google Colaboratory.

2. Материалы и методы

В данном исследовании используется Google Colaboratory для анализа данных.

Набор данных о ценах на автомобили в Австралии содержит информацию о различных типах автомобилей и их характеристиках. Он охватывает различные марки, модели, типы автомобилей, продаваемых на австралийском рынке. Это дает полезную информацию о тенденциях и факторах, влияющих на цены на автомобили в Австралии.

Данные для работы можно скачать по ссылке <https://www.kaggle.com/code/nelgiriyeewithana/an-introduction-to-australian-vehicle-prices/notebook>.

3. Результаты

Создадим блокнот в Google Colab и загрузим файл "Australian Vehicle Prices.csv".

В блоке кода импортируем необходимые библиотеки и модули для анализа данных, в данном случае Pandas, Plotly.express, Plotly.graph_objects. Загружаем данные из CSV-файла (Рис.1).

```
import plotly.express as px
import plotly.graph_objects as go
import pandas as pd

[2] df = pd.read_csv("Australian Vehicle Prices.csv")
```

Рис.1-Код

Далее распределяем цены на автомобили.

```
fig1 = px.histogram(df, x="Price", nbins=50, title="Distribution of Car Prices")
fig1.show()
```

Рис.2-Код

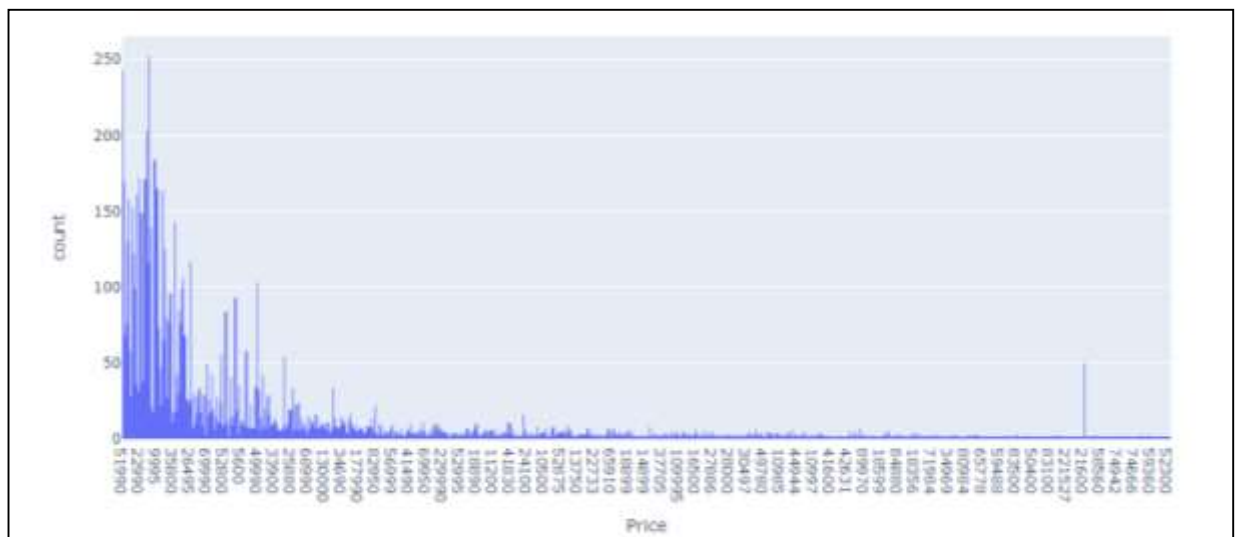


Рис.3-Итог распределения цен

Теперь пишем код для распределения по типам автомобилей.

```
fig2 = px.pie(df, names='Car/Suv', title='Distribution of Car Types')
fig2.show()
```

Рис.4-Код

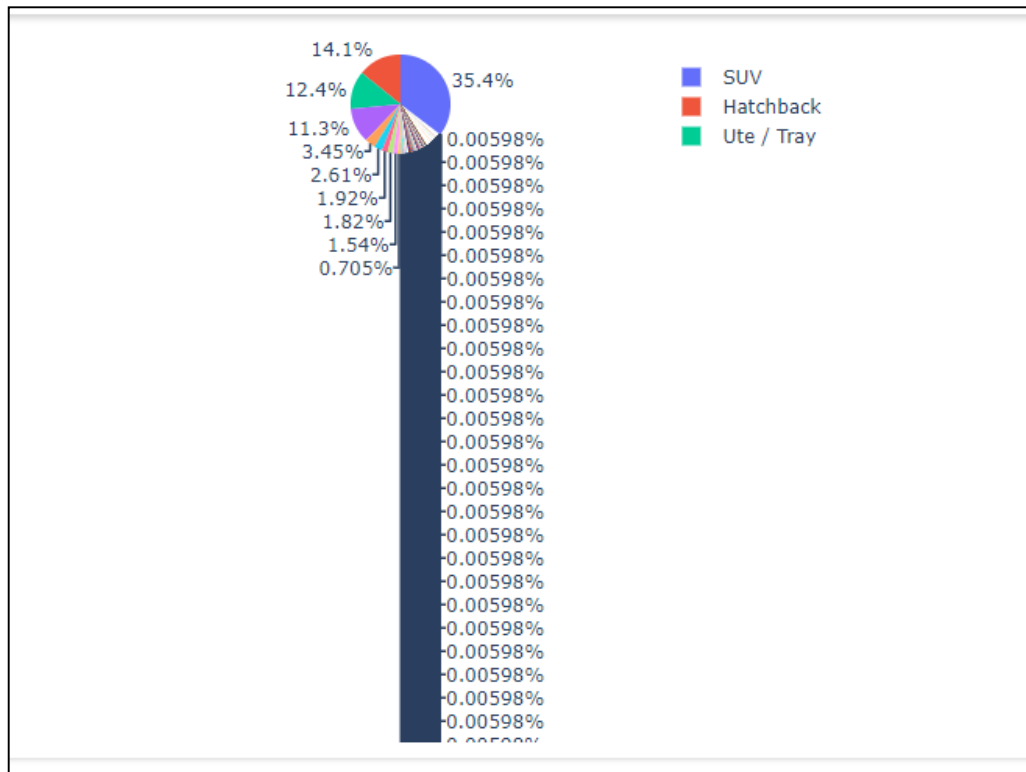


Рис.5-Итог распределения по типам автомобилей

Далее распределим по типам передач.

```
fig3 = px.pie(df, names='Transmission', title='Distribution of Transmission Types')
fig3.show()
```

Рис.6-Код

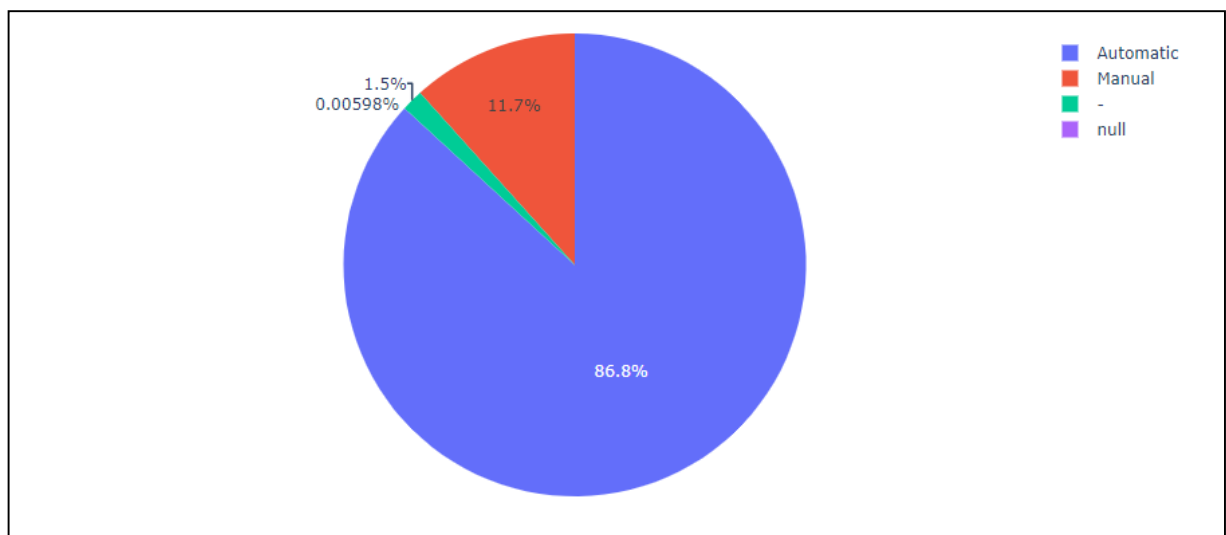


Рис.7-Итог по распределению передач

Теперь распределим по типам топлива.

```
fig4 = px.pie(df, names='FuelType', title='Distribution of Fuel Types')
fig4.show()
```

Рис.8-Код

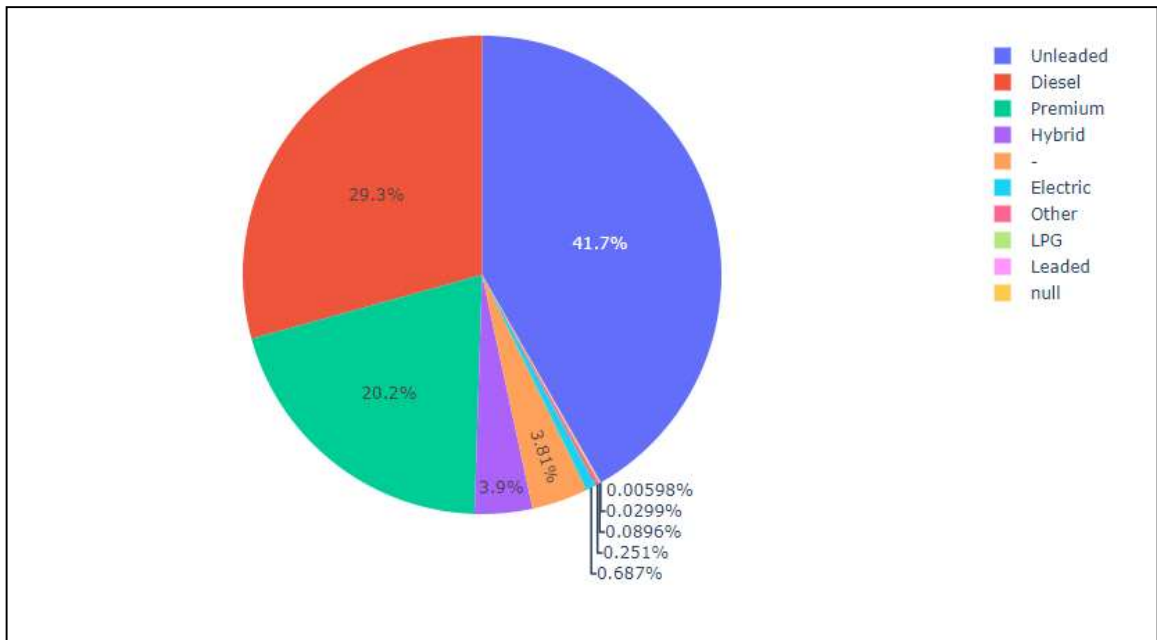


Рис.9-Итог распределения по топливу

Далее посмотрим соотношение расхода топлива и цены.

```
fig5 = px.scatter(df, x='FuelConsumption', y='Price', color='FuelType',
                  title='Fuel Consumption vs. Car Price', labels={'FuelConsumption': 'Fuel Consumption (L/100km)'})
fig5.show()
```

Рис.10-Код



Рис.11-Соотношение расхода топлива и цены

В данном случае, можно выбрать нужный вид топлива (Рис.12)



Рис.12-Вид топлива diesel

Далее смотрим на взаимосвязь между мощностью двигателя и ценой.

```
fig6 = px.scatter(df, x='Engine', y='Price', color='Brand',
                 title='Engine Capacity vs. Car Price', labels={'Engine': 'Engine Capacity (L or kW)'})
fig6.show()
```

Рис.13-Код



Рис.14-Итог

Таблица цен в разбивке по типу трансмиссии.

```
fig7 = px.box(df, x='Transmission', y='Price',
              title='Boxplot of Car Price by Transmission Type', labels={'Price': 'Car Price'})
fig7.show()
```

Рис.15-Код

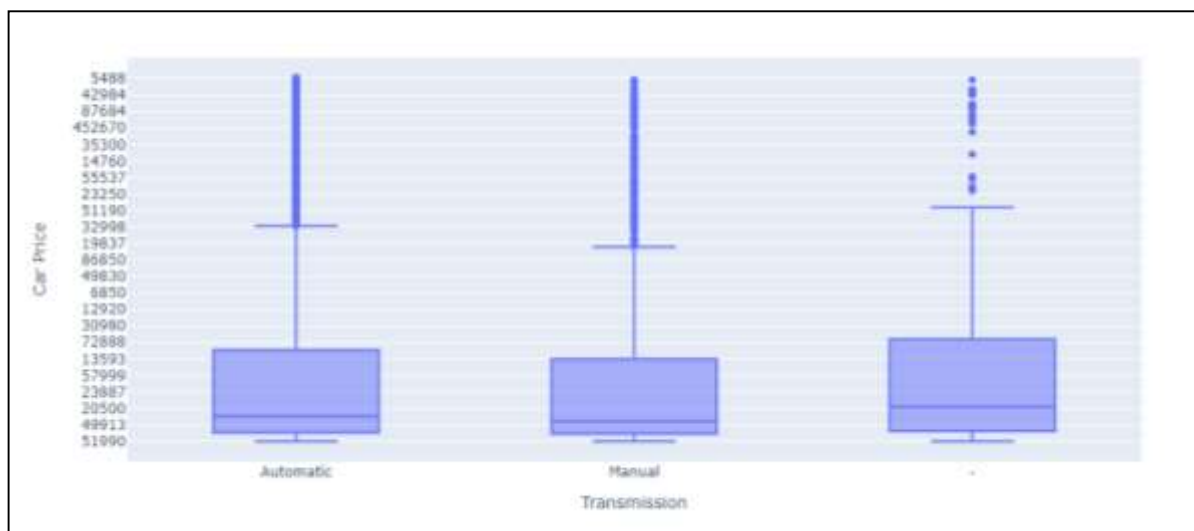


Рис.16-Итог трансмиссии

4. Выводы

В результате были выполнены следующие этапы: загрузка данных, ознакомление с данными, визуализация данных с использованием графиков и диаграмм.

Результаты исследования позволили выявить несколько ключевых факторов, оказывающих наибольшее влияние на ценообразование автомобилей в Австралии.

Исследование подтверждает эффективность использования Google Colaboratory для анализа данных в области автомобильной индустрии. Преимущества инструмента, такие как интеграция с сервисами Google, поддержка популярных библиотек и простота использования, делают его предпочтительным выбором для анализа данных и машинного обучения.

Библиографический список

1. Сухинина А.Р., Королева Е.В., Ефимов Е.А., Благой Н.А. Доступность жилья и определяющие ее факторы в регионах России. 2022. С. 70-81.
2. Косых Н. Е. Оценка гиперпараметров при анализе тональности русскоязычного корпуса текстов//Интеллектуальные технологии на транспорте. 2020. №. 3 (23). С. 41-44.
3. Богданов П. Ю. и др. Программные среды для изучения основ нейронных сетей //Программные продукты и системы. 2021. №. 1. С. 145-150.
4. Григорьев Е. А., Климов Н. С. Разведочный анализ данных с помощью python //E-Scio. 2020. №. 2 (41). С. 165-176.