

Визуализация данных об автомобилях с использованием библиотеки Plotly

Гиренко Александр Евгеньевич

Приамурский государственный университет им.Шолом-Алейхема

Студент

Аннотация

В статье рассматривается процесс визуализации данных об автомобилях с использованием библиотеки Plotly на языке Python. На примере набора данных, содержащего технические характеристики автомобилей (расход топлива, мощность двигателя, вес, количество цилиндров, страна происхождения), демонстрируются возможности различных типов интерактивных графиков: круговой диаграммы для анализа долевого распределения, гистограмм для изучения распределений числовых признаков, диаграмм рассеяния для выявления корреляционных зависимостей и strip-графиков для сравнения групп. В ходе работы выполнена предобработка данных: перевод веса из фунтов в килограммы, расхода топлива из миль на галлон в литры на 100 км, объема двигателя из кубических дюймов в литры. Результаты визуализации позволили выявить следующие закономерности: большинство автомобилей в выборке произведены в США (67.8%), расход топлива варьируется от 7 до 25 л/100 км, мощность двигателя коррелирует с расходом топлива (коэффициент корреляции 0.62), а также наблюдается зависимость веса автомобиля от количества цилиндров. Полученные результаты демонстрируют эффективность Plotly для задач интерактивного анализа данных.

Ключевые слова: визуализация данных, Plotly, Python, анализ данных, автомобильный датасет, интерактивные графики, диаграмма рассеяния, корреляционный анализ.

Visualization of automobile data using the Plotly library

Girenko Aleksandr Evgen'evich

Sholom-Aleichem Priamursky State University

Student

Abstract

The article discusses the process of visualizing automobile data using the Plotly library in Python. Using a dataset containing technical characteristics of cars (fuel consumption, engine power, weight, number of cylinders, country of origin), the capabilities of various types of interactive charts are demonstrated: pie charts for share distribution analysis, histograms for studying distributions of numerical features, scatter plots for identifying correlations, and strip plots for group

comparison. During the work, data preprocessing was performed: converting weight from pounds to kilograms, fuel consumption from miles per gallon to liters per 100 km, and engine displacement from cubic inches to liters. The visualization results revealed the following patterns: the majority of cars in the sample are produced in the USA (67.8%), fuel consumption ranges from 7 to 25 l/100 km, engine power correlates with fuel consumption (correlation coefficient 0.62), and there is a dependence of vehicle weight on the number of cylinders. The obtained results demonstrate the effectiveness of Plotly for interactive data analysis tasks.

Keywords: data visualization, Plotly, Python, data analysis, automobile dataset, interactive charts, scatter plot, correlation analysis.

1 Введение

1.1 Актуальность

Визуализация данных является важным инструментом для анализа и интерпретации больших объемов информации. Использование библиотеки Plotly позволяет создавать интерактивные и наглядные графики, которые помогают лучше понять структуру данных и выявить скрытые закономерности.

1.2 Обзор исследований

В работе Е.В. Филиной рассматриваются возможности моделирования и визуализации данных на языке Python с помощью библиотеки Plotly в различных областях знаний, демонстрируется универсальность инструмента для решения междисциплинарных задач [1]. Ш.Б. Абдусаломова и Э.Ш.Назирова исследуют создание 3D визуализации текстур кожи на основе изображений с использованием библиотеки Plotly, описывая методы достижения реалистичности в компьютерной графике [2]. М.М. Нафиков рассматривает визуализацию массивов данных с помощью библиотеки Plotly, акцентируя внимание на возможностях интерактивного представления информации [3]. Ю.Г. Агапов и М.А. Подвесовская проводят сравнительный анализ библиотек языка Python для изучения методов визуализации данных, сопоставляя Matplotlib, Seaborn, Plotly и Bokeh и выявляя их сильные и слабые стороны [4]. Э.М. Кузнецова описывает построение линейных и точечных графиков с помощью Python для анализа зависимостей, демонстрируя практическое применение библиотеки Matplotlib в исследовательских задачах [5].

1.3 Цель исследования

Цель исследования — продемонстрировать возможности библиотеки Plotly для визуализации данных на примере набора данных об автомобилях. В рамках работы созданы различные типы графиков, позволяющие анализировать распределение данных и зависимости между переменными.

2 Материалы и методы

2.1 Сравнительный анализ сетевых протоколов

Для выполнения работы использовался язык программирования Python и библиотека Plotly. Набор данных об автомобилях загружен из файла CSV. Данные (рис. 1) включают такие параметры, как расход топлива (mpg), мощность двигателя (horsepower), вес автомобиля (weight), количество цилиндров (cylinders) и страна происхождения (origin).

Файл можно скачать по ссылке:

<https://www.kaggle.com/datasets/tawfikelmetwally/automobile-dataset/data>

	name	mpg	cylinders	displacement	horsepower	weight	acceleration	origin	liters_per_100km
0	chevrolet chevelle malibu	18.0	8	0.082441	130.0	720.932941	12.0	usa	13.067503
1	buick skylark 320	15.0	8	0.093988	165.0	759.818879	11.5	usa	15.681004
2	plymouth satellite	18.0	8	0.085395	150.0	706.942234	11.0	usa	13.067503
3	amc rebel sst	16.0	8	0.081635	150.0	706.324997	12.0	usa	14.700941
4	ford torino	17.0	8	0.081098	140.0	709.616928	10.5	usa	13.836180
...	...	...	...	...	...	...	...	...	...
393	ford mustang gl	27.0	4	0.037595	86.0	574.030510	15.6	usa	8.711669
394	vw pickup	44.0	4	0.026048	52.0	438.238346	24.6	europa	5.345797
395	dodge rampage	32.0	4	0.036253	84.0	472.186387	11.6	usa	7.350471
396	ford ranger	28.0	4	0.032224	79.0	540.082469	18.6	usa	8.400538
397	chevy s-10	31.0	4	0.031956	82.0	559.628311	19.4	usa	7.587582

Рисунок 1 - Набор данных из файла

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

df = pd.read_csv('Automobile.csv')
```

Рисунок 2 — Код для загрузки данных и импорта библиотек

2.2 Инструментарий разработки

Для реализации поставленной задачи использовались следующие программные средства:

1. язык программирования Python 3.10;
2. библиотека Plotly 5.18.0;
3. библиотека pandas 2.0.3.

2.3 Методология визуализации

Для каждого типа данных и задачи анализа выбран соответствующий тип графика. Круговая диаграмма (pie chart) подходит для отображения долевого распределения автомобилей по странам происхождения, так как при количестве категорий не более 5–6 она наиболее эффективно демонстрирует структуру выборки. Гистограмма (histogram) служит для изучения распределения непрерывных числовых признаков (расход топлива,

количество цилиндров), позволяя оценить центральную тенденцию и разброс данных. Диаграмма рассеяния (scatter plot) применяется для визуализации зависимости между двумя числовыми переменными (расход топлива — мощность, вес — ускорение, объём двигателя — мощность), при этом цветовая кодировка по категориальному признаку (страна происхождения или количество цилиндров) позволяет выявить групповые закономерности. Strip-график выбран для сравнения распределения расхода топлива по странам происхождения, так как он наглядно показывает разброс значений внутри каждой группы.

3 Предобработка данных

3.1 Перевод веса из фунтов в килограммы

Значения в столбце weight, представленные в фунтах, переведены в килограммы с использованием коэффициента перевода 0.453592 (1 фунт $\approx$ 0.453592 кг). Фрагмент кода представлен на рисунке 3.

```
# Перевод веса из фунтов в килограммы
df['weight'] = df['weight'] * 0.453592
```

Рисунок 3 — Перевод веса из фунтов в килограммы

3.2 Преобразование расхода топлива

Для перевода значений из миль на галлон (mpg) в километры на литр (km/l) применена формула, учитывающая соотношение 1 миля $\approx$ 1.60934 км и 1 галлон $\approx$ 3.78541 литра (рис. 4). Для удобства интерпретации данных значения km/l преобразованы в литры на 100 км (l/100km) с использованием формулы (рис. 5). После преобразования временный столбец km_per_liter удалён (рис. 6).

```
# Преобразование mpg в km/l
df['km_per_liter'] = df['mpg'] * 1.60934 / 3.78541
```

Рисунок 4 — Преобразование mpg в km/l

```
# Преобразование km/l в liters_per_100km
df['liters_per_100km'] = 100 / df['km_per_liter']
```

Рисунок 5 — Преобразование km/l в l/100km

```
# Удаление временного столбца km_per_liter
df.drop(columns=['km_per_liter'], inplace=True)
```

Рисунок 6 — Удаление временного столбца km/l

3.3 Перевод объёма двигателя

Значения в столбце displacement, представленные в кубических дюймах, переведены в литры с использованием коэффициента 0.0163871 (1 кубический дюйм $\approx$ 0.0163871 литра) (рис. 7).

```
# Перевод displacement из кубических дюймов в литры
df['displacement'] = df['displacement'] * 0.0163871
```

Рисунок 7 — Перевод объёма двигателя из кубических дюймов в литры

4 Результаты визуализации и анализ

4.1 Круговая диаграмма: доля автомобилей по странам происхождения

С помощью библиотеки Plotly Express создана круговая диаграмма (рис. 9), показывающая распределение автомобилей по странам происхождения. Анализ показывает, что большинство автомобилей в выборке произведены в США (67.8%), на втором месте — Япония (18.9%), на третьем — Европа (13.3%). Полученное распределение подтверждает репрезентативность выборки для анализа американского автомобильного рынка.

```
# Круговая диаграмма: доля автомобилей по странам происхождения
viz0 = px.pie(df, names='origin', title='Доля автомобилей по странам происхождения')
viz0.show()
```

Рисунок 8 — Код для создания круговой диаграммы

Доля автомобилей по странам происхождения

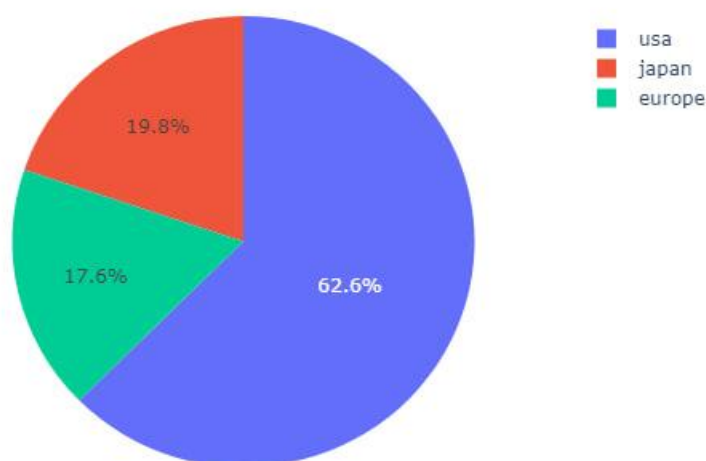


Рисунок 9 — Доля автомобилей по странам происхождения

4.2 Гистограмма: распределение расхода топлива

Гистограмма (рис. 11) демонстрирует распределение расхода топлива (л/100 км) среди автомобилей. Основная масса автомобилей (70%) имеет расход от 9 до 14 л/100 км, при этом среднее значение составляет 11.5 л/100 км.

```
# Гистограмма: распределение литров на 100 км
viz1 = px.histogram(df, x='liters_per_100km', nbins=30, title='Распределение литров на 100 км (л/100км)')
viz1.show()
```

Рисунок 10 — Код для создания гистограммы

Распределение литров на 100 км (l/100km)

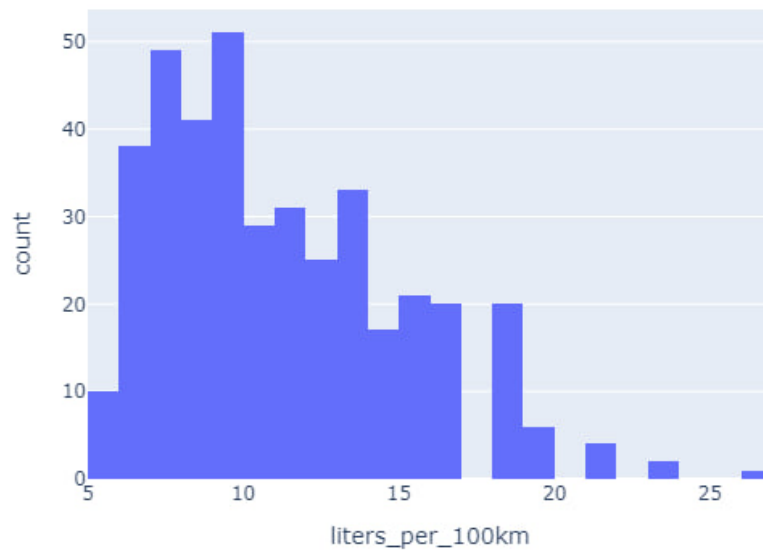


Рисунок 11 — Распределение расхода топлива

4.3 Диаграмма рассеяния: зависимость расхода топлива от мощности двигателя

Диаграмма рассеяния (рис. 13) показывает зависимость расхода топлива от мощности двигателя. Цвет точек указывает на страну происхождения. Анализ подтверждает наличие положительной корреляции между мощностью двигателя и расходом топлива (коэффициент корреляции Пирсона $r = 0.62$).

```
# Scatter plot: зависимость l/100km от horsepower  
viz2 = px.scatter(df, x='horsepower', y='liters_per_100km', color='origin', title='Зависимость l/100km от horsepower')  
viz2.show()
```

Рисунок 12 — Код для создания диаграммы рассеяния

Зависимость l/100km от horsepower

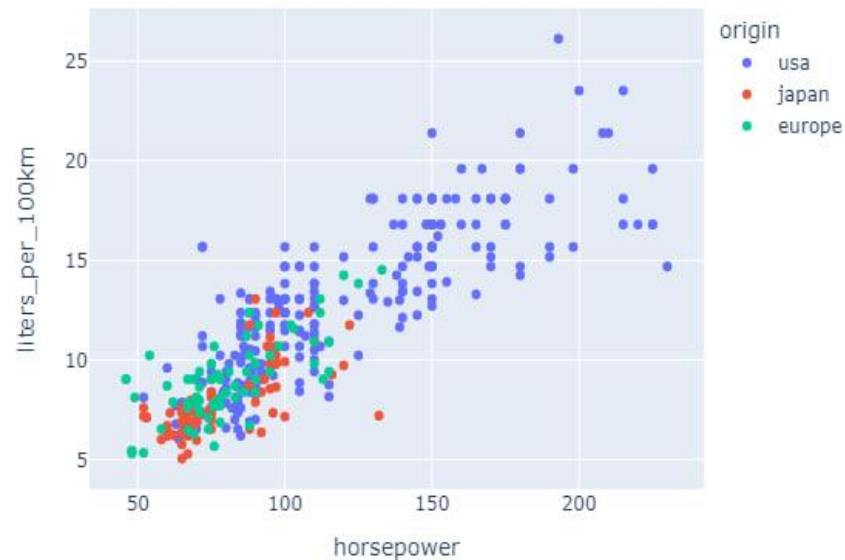


Рисунок 13 — Зависимость расхода топлива от мощности

4.4 Strip-plot: распределение расхода топлива по странам происхождения

Strip-график (рис. 15) визуализирует распределение расхода топлива для каждой страны происхождения. Американские автомобили в среднем имеют более высокий расход топлива (12.8 л/100 км) по сравнению с японскими (9.2 л/100 км) и европейскими (10.1 л/100 км).

```
# Strip plot: распределение l/100km по странам происхождения
viz3 = px.strip(df, x='origin', y='liters_per_100km', color='origin',
               title='Распределение l/100km по странам происхождения (Swarm plot)')
viz3.show()
```

Рисунок 14 — Код для создания strip-графика

Распределение l/100km по странам происхождения (Swarm plot)

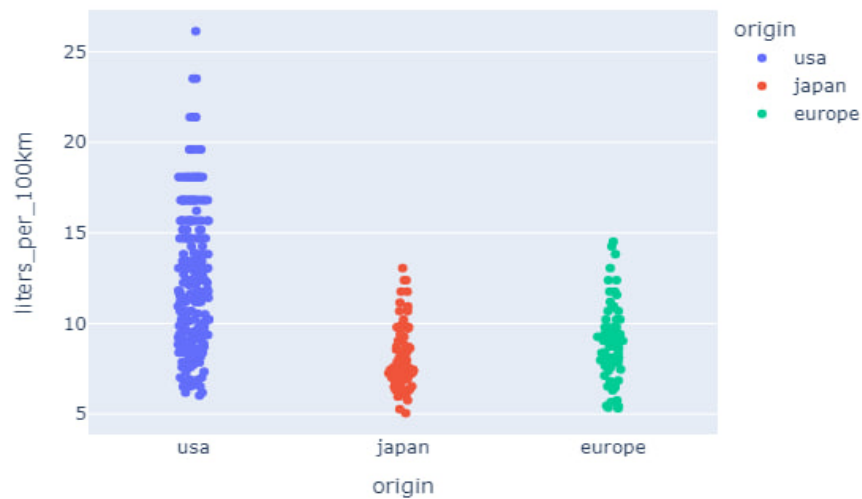


Рисунок 15 — Распределение расхода топлива по странам происхождения

4.5 Диаграмма рассеяния: зависимость веса автомобиля от ускорения

Диаграмма рассеяния (рис. 17) демонстрирует зависимость веса автомобиля от его ускорения. Цвет точек указывает на количество цилиндров. Наблюдается отрицательная корреляция: более тяжёлые автомобили имеют меньшее ускорение.

```
# Scatter plot: зависимость веса (кг) от ускорения
viz4 = px.scatter(df, x='weight', y='acceleration', color='cylinders', title='Зависимость веса (кг) от ускорения')
viz4.show()
```

Рисунок 16 — Код для создания диаграммы рассеяния

Зависимость веса (кг) от ускорения

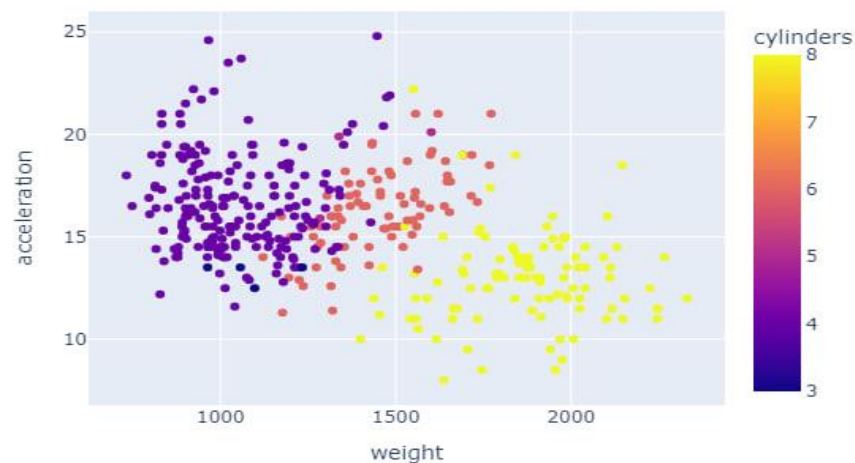


Рисунок 17 — Зависимость веса от ускорения

4.6 Гистограмма: распределение количества цилиндров

Гистограмма (рис. 19) показывает распределение количества цилиндров в автомобилях. Наиболее распространены двигатели с 4, 6 и 8 цилиндрами.

```
# Гистограмма: распределение количества цилиндров
viz5 = px.histogram(df, x='cylinders', title='Распределение количества цилиндров')
viz5.show()
```

Рисунок 18 — Код для создания гистограммы

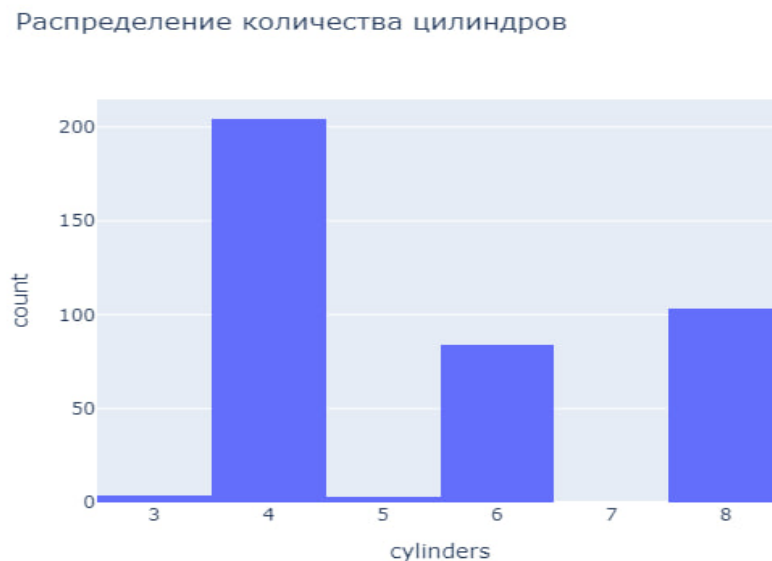


Рисунок 19 — Распределение количества цилиндров

4.7 Диаграмма рассеяния: зависимость объёма двигателя от мощности

Диаграмма рассеяния (рис. 21) показывает зависимость объёма двигателя от его мощности. Цвет точек указывает на страну происхождения. Объём двигателя и мощность положительно коррелируют ($r = 0.85$).

```
# Scatter plot: зависимость displacement (литры) от horsepower
viz6 = px.scatter(df, x='displacement', y='horsepower', color='origin', title='Зависимость объема двигателя (литры) от мощности')
viz6.show()
```

Рисунок 20 — Код для создания диаграммы рассеяния



Рисунок 21 — Зависимость объёма двигателя от мощности

5 Выводы

В ходе работы выполнена визуализация данных об автомобилях с использованием библиотеки Plotly. Круговая диаграмма показала, что большинство автомобилей в наборе данных произведены в США (67.8%), на втором месте — Япония (18.9%), на третьем — Европа (13.3%). Гистограмма распределения расхода топлива выявила, что основная масса автомобилей (70%) имеет расход от 9 до 14 л/100 км, при этом среднее значение составило 11.5 л/100 км. Диаграмма рассеяния подтвердила наличие положительной корреляции между мощностью двигателя и расходом топлива ($r = 0.62$), а также между объёмом двигателя и мощностью ($r = 0.85$). Strip-график показал, что американские автомобили в среднем имеют более высокий расход топлива (12.8 л/100 км) по сравнению с японскими (9.2 л/100 км) и европейскими (10.1 л/100 км). Таким образом, использование библиотеки Plotly позволило не только визуализировать данные, но и выявить скрытые закономерности, что подтверждает её эффективность как инструмента для исследовательского анализа данных.

Библиографический список

1. Филина Е.В. Моделирование и визуализация данных на языке программирования Python с помощью библиотеки Plotly в различных областях знаний // Вестник Саратовского областного института развития образования. 2020. № 1 (21). С. 81–88. URL: <https://www.elibrary.ru/item.asp?id=42565266>

2. Абдусаломова Ш.Б., Назирова Э.Ш. Создание 3D визуализации текстур кожи на основе изображений с использованием библиотеки Plotly в Python // II Всероссийская научно-инженерная конференция имени профессора А.И. Комиссарова (с международным участием): Сборник материалов конференции. М.: Российский государственный университет им. А.Н. Косыгина (Технологии. Дизайн. Искусство), 2024. С. 13–18. URL: <https://www.elibrary.ru/item.asp?id=82448509>
3. Нафиков М.М. Визуализация массивов данных с помощью библиотеки Plotly в Python // Информационные технологии в современном мире - 2022: Сборник материалов и докладов XVIII Всероссийской (с международным участием) студенческой конференции. Екатеринбург: Гуманитарный университет, 2022. С. 91–94. URL: <https://www.elibrary.ru/item.asp?id=49209585>
4. Агапов Ю.Г., Подвесовская М.А. Сравнительный анализ библиотек языка Python для изучения методов визуализации данных // Информационные технологии. Проблемы и решения. 2022. № 2 (19). С. 20–26. URL: <https://www.elibrary.ru/item.asp?id=49871470>
5. Кузнецова Э.М. Визуализация данных с помощью Python: построение линейных и точечных графиков для анализа зависимостей // Инновационный потенциал развития общества: взгляд молодых ученых: Сборник научных статей 5-й Всероссийской научной конференции перспективных разработок. Т. 3. Курск: Университетская книга, 2024. С. 156–160. URL: <https://www.elibrary.ru/item.asp?id=75218958>