

Анализ задержек авиарейсов с использованием библиотеки Altair

Анишкова Анастасия Сергеевна

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

В работе реализовано применение библиотеки Altair для построения интерактивных и статических визуализаций данных на примере датасета о задержках авиарейсов в облачной среде Google Colab. Среда Google Colab обеспечила удобную и доступную платформу для выполнения вычислений без необходимости локальной настройки окружения. Работа может быть использована в образовательных целях для изучения современных методов визуализации данных в Python.

Ключевые слова: Altair, Google Colab, Python, Анализ задержек авиарейсов.

Analyzing flight delays using the Altair library

Anishkova Anastasia Sergeevna

Sholom-Aleichem Priamurskiy State University

Student

Abstract

The paper implements the use of the Altair library to build interactive and static data visualizations using the dataset of flight delays in the Google Colab cloud environment. The Google Colab environment provided a convenient and accessible platform for performing calculations without the need for local environment configuration. The work can be used for educational purposes to learn modern data visualization methods in Python.

Key words: Altair, Google Colab, Python, and Flight Delay Analysis.

1 Введение

1.1 Актуальность

В современном мире анализ и визуализация данных играют ключевую роль в принятии обоснованных решений в различных сферах деятельности — от бизнеса и экономики до науки и образования. С ростом объемов данных возрастает потребность в удобных и эффективных инструментах для их обработки и наглядного представления. Библиотеки визуализации данных, такие как Altair, позволяют создавать как статические, так и интерактивные графики, основываясь на принципах грамматики графики, что делает их мощным и интуитивно понятным инструментом для анализа.

1.2 Обзор исследований

П. А. Пылов, А. В. Протоद्याконов описали программные библиотеки языка программирования Python для реализации алгоритмов визуализации данных [1]. Сравнительный анализ библиотек визуализации данных для задачи обработки образовательной информации провели О. А. Шутова, Д. А. Грамаков [2]. А. Оразгулыев, С. А. Гараджаева визуализировали данные и представили результаты исследования [3]. Библиотеки Python для визуализации научных концепций в образовательных целях применили Н. К. Аркабаев, Б. У. Доолотбек [4]. Е. В. Филина смоделировала и визуализировала данные на языке программирования Python с помощью библиотеки Plotly в различных областях знаний [5]

1.3 Цель исследования

Разработать и визуализировать анализ данных о задержках авиарейсов с использованием библиотеки Altair в облачной среде Google Colab, демонстрируя возможности инструмента для построения интерактивных и информативных графиков.

2 Материалы и методы

Использовались библиотека Altair для анализа данных о задержках авиарейсов.

3 Результаты

Ссылка на ноутбук:
https://colab.research.google.com/drive/1btKOoaHqg4KqRPzvz2iWRATo0w9GI_ZR?usp=sharing.

Для работы с визуализацией данных в Python установим библиотеку Altair, предназначенную для создания графиков на основе грамматики графики, и библиотеку `vega_datasets`, предоставляющую набор готовых данных для построения примеров (рис.1).

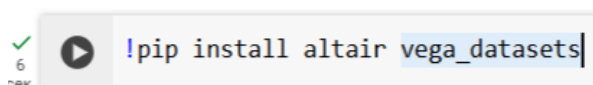


Рисунок 1. Установка библиотек

Для обработки данных и построения визуализаций подключим библиотеки Altair, предназначенную для создания графиков и диаграмм, и pandas (сокращённо `pd`), используемую для работы с табличными данными в формате DataFrame (рис.2).

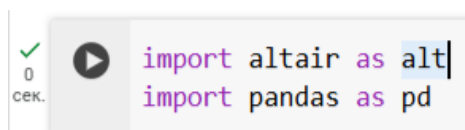


Рисунок 2. Импорт библиотек

Загрузим встроенный набор данных о пассажирских рейсах `flights_5k` из библиотеки `vega_datasets`, после чего с помощью `Altair` построим интерактивный график, на котором по оси X отложена дата рейса, по оси Y — задержка, цвет точек зависит от аэропорта вылета, а при наведении курсора отображаются дополнительные сведения; график имеет заголовок «Задержки рейсов по аэропортам» и поддерживает масштабирование и прокрутку (рис.3).

Используем встроенный набор данных с пассажиропотоком

```
from vega_datasets import data
```

```
flights = data.flights_5k()
```

Построим график задержек рейсов

```
chart = alt.Chart(flights).mark_circle().encode(
    x='date',
    y='delay',
    color='origin',
    tooltip=['date', 'delay', 'destination', 'origin']
).properties(
    title='Задержки рейсов по аэропортам'
).interactive()
```

chart



Рисунок 3. График

Создадим гистограмму распределения задержек рейсов с помощью библиотеки Altair, где по оси X откладываются интервалы задержек, а по оси Y — количество рейсов, попадающих в каждый интервал; график отображает форму распределения задержек и имеет заголовок «Распределение задержек рейсов» (рис.4).

```
# Гистограмма распределения задержек
histogram = alt.Chart(flights).mark_bar().encode(
    alt.X("delay:Q", bin=alt.Bin(maxbins=30)),
    y='count()'
).properties(
    title='Распределение задержек рейсов'
)
```

histogram

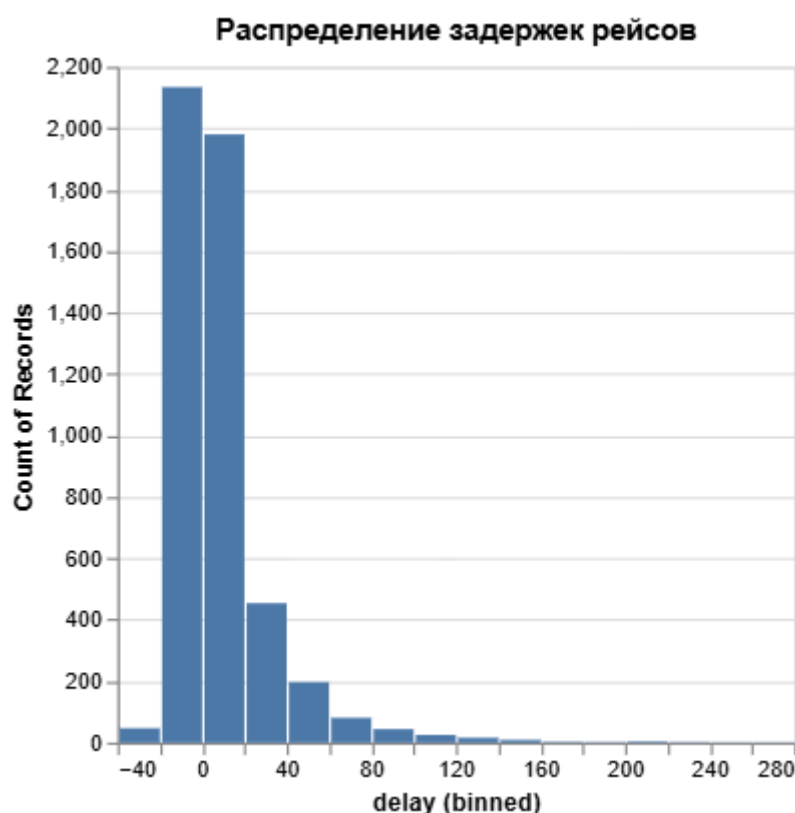


Рисунок 4. Гистограмма

Построим линейный график, отображающий среднюю задержку рейсов по дням, где по оси X указаны даты, а по оси Y — значения средней задержки; при наведении курсора отображаются конкретные значения даты и средней задержки, а сам график имеет заголовок «Средняя задержка рейсов по дням» (рис.5).

```
# Линейный график: средняя задержка по дням
line_chart = alt.Chart(flights).mark_line().encode(
    x=alt.X('date:T', title='Дата'),
    y=alt.Y('mean(delay):Q', title='Средняя задержка'),
    tooltip=['date', 'mean(delay)']
)
```

```

).properties(
  title='Средняя задержка рейсов по дням'
)

```

```

line_chart

```



Рисунок 5. Линейный график

Создадим столбчатую диаграмму для сравнения средней задержки рейсов по аэропортам вылета, где по оси X указаны названия аэропортов, а по оси Y — значения средней задержки; столбцы окрашены в разные цвета в зависимости от аэропорта, а при наведении курсора отображаются точные значения средней задержки для каждого из них; график имеет заголовок «Средняя задержка по аэропортам вылета» (рис.6).

Столбчатая диаграмма: сравнение средней задержки по аэропортам

```

bar_chart = alt.Chart(flights).mark_bar().encode(
  x=alt.X('origin:N', title='Аэропорт вылета'),
  y=alt.Y('mean(delay):Q', title='Средняя задержка'),
  color='origin:N',
  tooltip=['origin', 'mean(delay)']
).properties(
  title='Средняя задержка по аэропортам вылета'
)

```

```

bar_chart

```

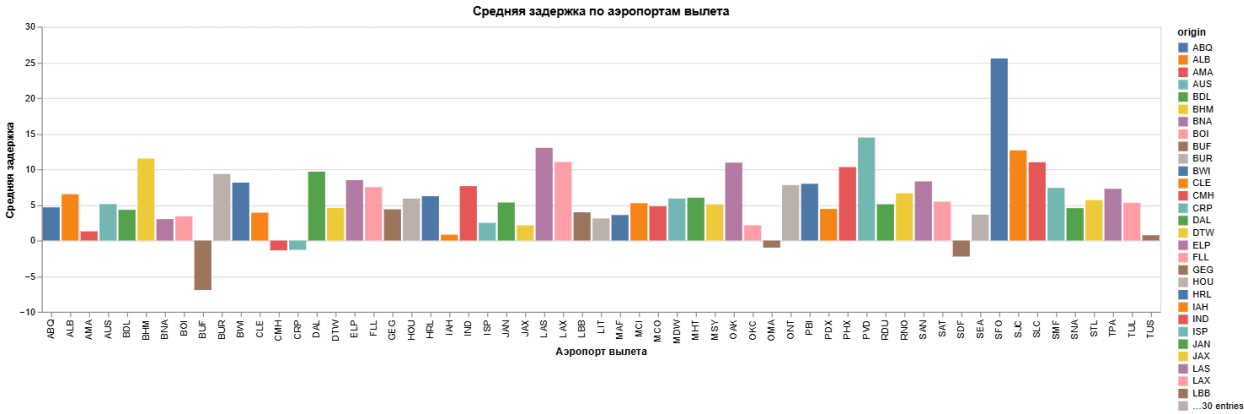


Рисунок 6. Столбчатая диаграмма

Построим столбчатая диаграмма, отображающая количество рейсов по пунктам назначения, где по оси X указано число рейсов, а по оси Y — названия пунктов назначения, отсортированные по убыванию; столбцы окрашены в разные цвета для каждого пункта, а при наведении курсора отображаются точные значения; график имеет заголовок «Количество рейсов по пунктам назначения» (рис.7).

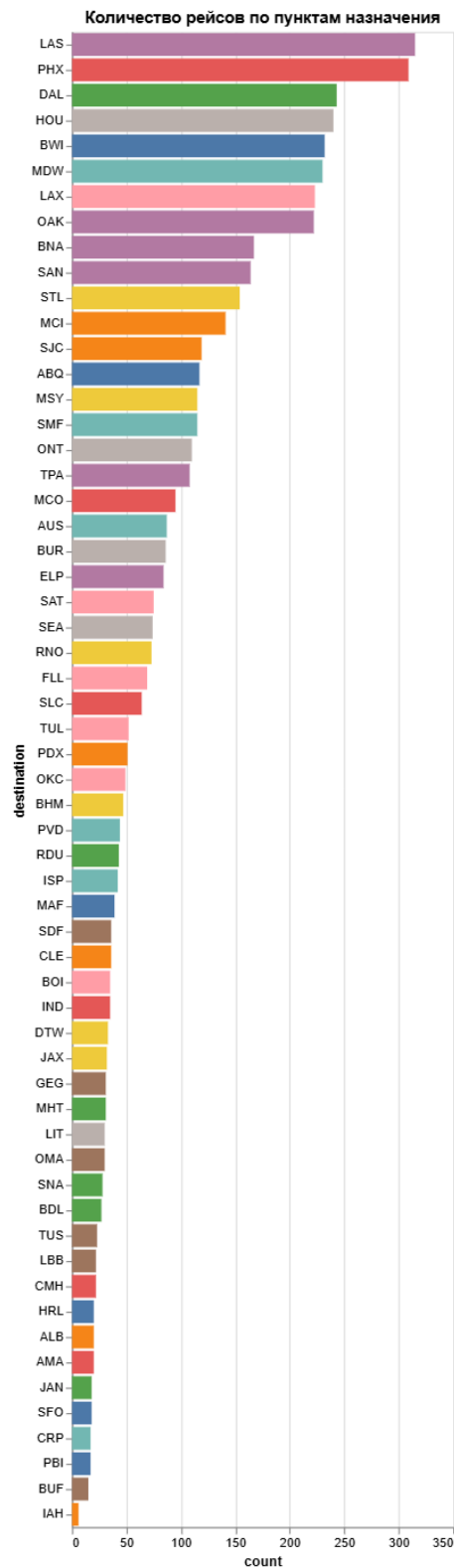


Рисунок 7. Столбчатая диаграмма

Создадим столбчатую диаграмму (рис.8), отображающая топ-10 дней с наибольшим числом задержек рейсов свыше 60 минут, где по оси X указаны даты, а по оси Y — количество задержанных рейсов; столбцы окрашены в оранжевый цвет, а при наведении курсора отображаются точные значения даты и количества задержек; график имеет заголовок «Топ-10 дней с наибольшим числом задержек (> 60 мин)».

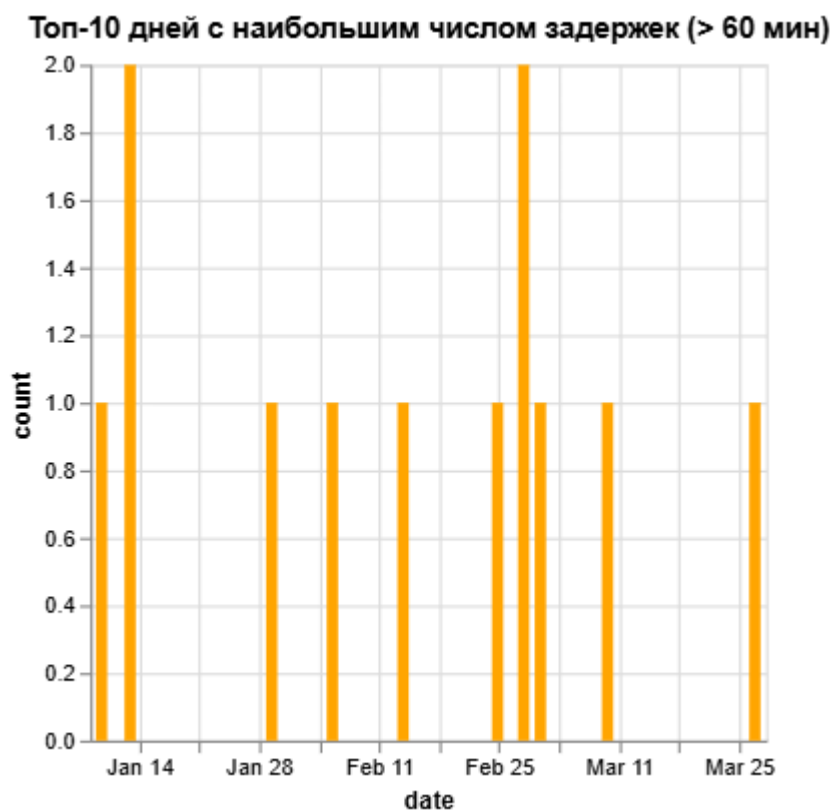


Рисунок 8. График

4. Выводы

Каждый график наглядно отображает различные аспекты данных, что позволяет выявлять закономерности, аномалии и тенденции в пассажирских перевозках. Визуализации выполнены в едином стиле, имеют заголовки, подписи осей и интерактивные элементы, повышающие удобство анализа. Таким образом, проделанная работа демонстрирует эффективность использования Altair для анализа и представления данных в понятной и информативной форме.

Библиографический список

1. Пылов П. А., Протоdjяконов А. В. Программные библиотеки языка программирования Python для реализации алгоритмов визуализации данных //Инновации. Наука. Образование. 2020. №. 23. С. 233-239.
2. Шутова О. А., Грамаков Д. А. Сравнительный анализ библиотек визуализации данных для задачи обработки образовательной информации //Проблемы теории и практики инновационного развития и интеграции современной науки и образования. 2021. С. 190-194.
3. Оразгулыев А., Гараджаева С. А. Визуализация данных. Представление результатов исследования //Наука и мировоззрение. 2025. Т. 1. №. 40. С. 42-48.
4. Аркабаев Н. К., Доолотбек Б. У. Применение библиотек python для визуализации научных концепций в образовательных целях //Science and innovation. 2024. Т. 3. №. Special Issue 50. С. 398-406.

5. 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. С. 81-88.