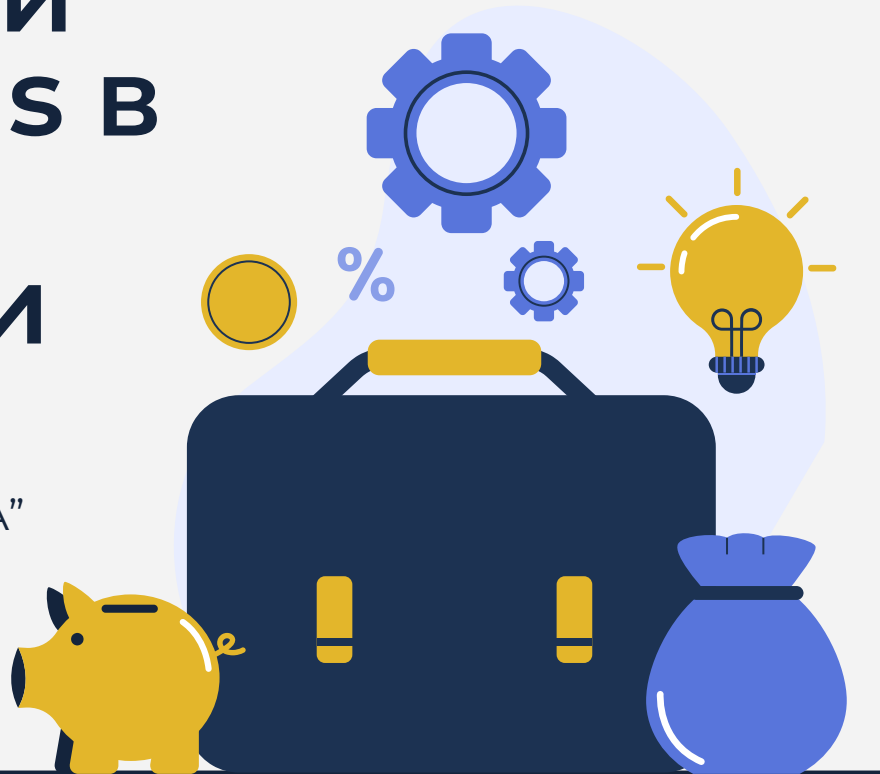


Сравнение возможностей Excel и Pandas в вопросах визуализации данных

Проект подготовлен учениками 10"А"
класса Жуковенем Сергеем и
Новиковым Вячеславом



Содержание

01

Актуальность темы

02

Цель и задачи

03

Проблема и
проведение опроса

04

Программы и языки

05

Визуализация данных
и аналитика

06

Итоги сравнения

Актуальность темы



Для чего все это?

В современном мире объемы данных постоянно увеличиваются, и эффективное их анализирование становится ключевой задачей для многих предприятий и исследовательских организаций. Данные подразделяются на два типа: персональные и промышленные



Рабочий вопрос

Актуальность данного исследования подчеркивается тем, что многие специалисты по всему миру сталкиваются с вопросом выбора между Excel и Pandas при работе с данными, и важно понимать, какие задачи каждый из них решает наилучшим образом.



Цель и задачи

02

Целью данного исследования является исследование и сравнение возможностей Excel и Pandas в решении задач обработки данных и визуализации.



Формулировка
статистических гипотез к
имеющимся данным.



Проработка
аналитических решений и
поиск оптимальных
методов анализа в
зависимости от специфики
данных и выбранных
гипотез.



Реализация алгоритмов
обработки и визуализации
данных с помощью Pandas.

Проблема

Какие инструменты, Excel или Pandas, предоставляют более эффективные средства для обработки и визуализации данных, учитывая различные сценарии использования?

Правда ли использование Pandas для анализа данных может обеспечить более высокую производительность и гибкость по сравнению с Excel, особенно при работе с большими объемами данных и сложными запросами.

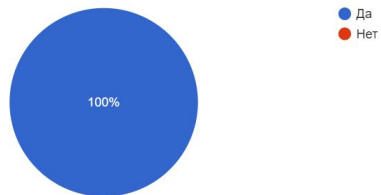
03



Ведение опроса

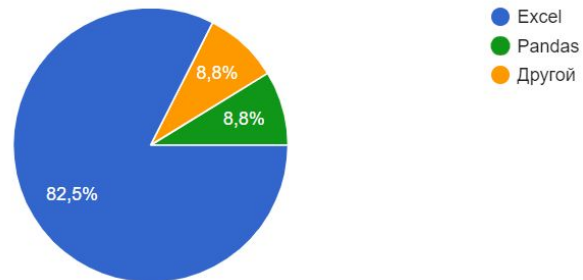
Знаете ли вы что такое Excel?

57 ответов



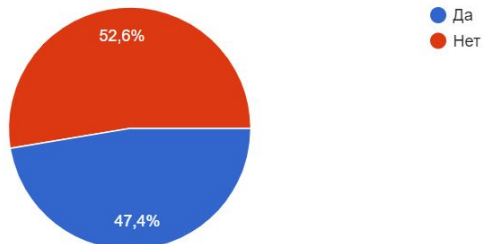
Какой инструмент вам кажется наиболее удобным?

57 ответов



Знаете ли вы что такое Pandas?

57 ответов



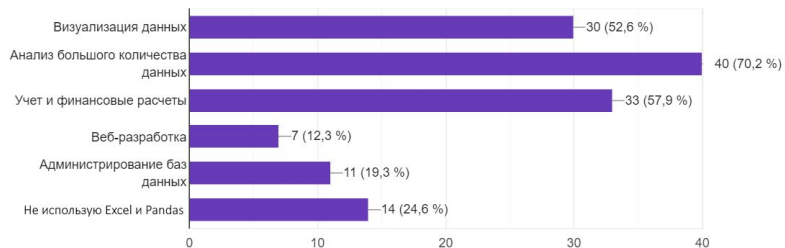
Какой можно сделать вывод?



Потребности опрашиваемых

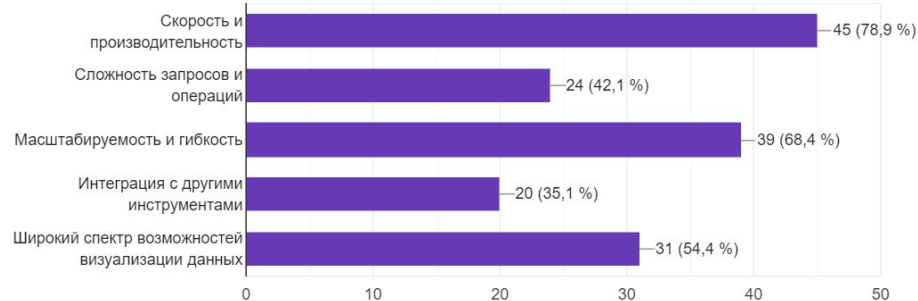
При выполнении каких задач вы используете Excel/Pandas?

57 ответов



Какие качества инструмента для вас наиболее важны?

57 ответов



Программы и языки

04

Excel

1. Мы использовали Excel - мощный инструмент для работы с электронными таблицами, предоставляющий широкий набор функций для обработки данных. В Excel, данные организованы в ячейках, что облегчает их структурирование. Он предоставляет пользователю возможность выполнять различные вычисления и анализировать данные с использованием формул и функций. Графические инструменты в Excel, такие как диаграммы и графики, позволяют визуализировать данные, что улучшает их восприятие. Интерактивные элементы, такие как сводные таблицы и диаграммы, делают Excel удобным инструментом для анализа данных различной сложности.



Pandas

2. Также была использована Pandas - библиотека Python, основанная на структурах данных, которая предоставляет мощные инструменты для обработки и анализа данных, особенно в формате таблицы. Она представляет собой инструмент для работы с данными, аналогичный таблицам и базам данных, но с возможностью использования языка программирования Python для выполнения различных операций.



Категории сравнения инструментов



Визуализация данных



Скорость и
производительность



Масштабируемость и
гибкость



Сложность запросов и
операций



Интеграция с другими
инструментами

Какие данные для сравнения
были использованы?

Результаты
опроса



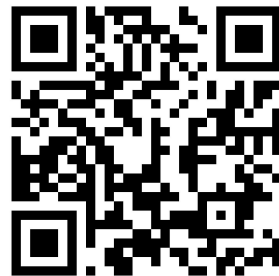
Большое количество
данных из открытого
доступа

Продукт

Продуктом проекта является методически-практическая подборка: репозиторий, содержащий два основных файла: Excel_London_crowding.xlsx и London_crowding.ipynb

Оба файла представляют собой методические отчеты о создании и реализации пайплайнов обработки и визуализации открытых данных (данных о нагрузке транспортных сетей Лондона за 2022 год).

Также репозиторий содержит директорию data с исходной базой данных в разных форматах файла.



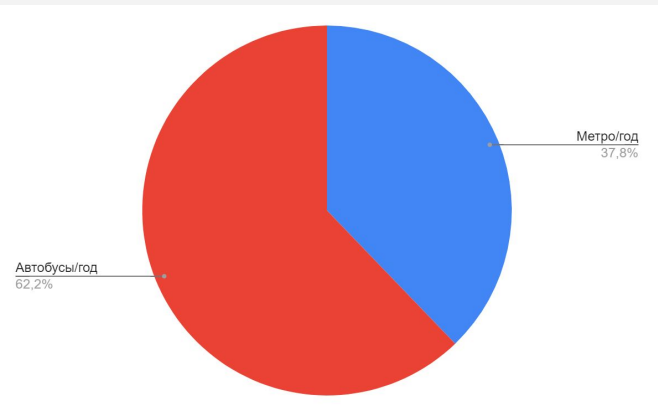
Визуализация данных и аналитика

05

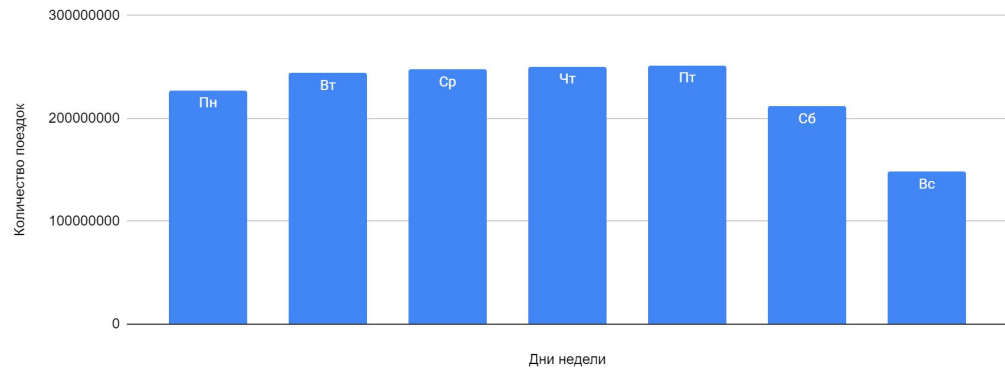
Pandas **не имеет** собственной визуализации

Excel предоставляет пользователю широкий выбор инструментов для удобства графического восприятия данных, в том числе - диаграммы, графики, карты

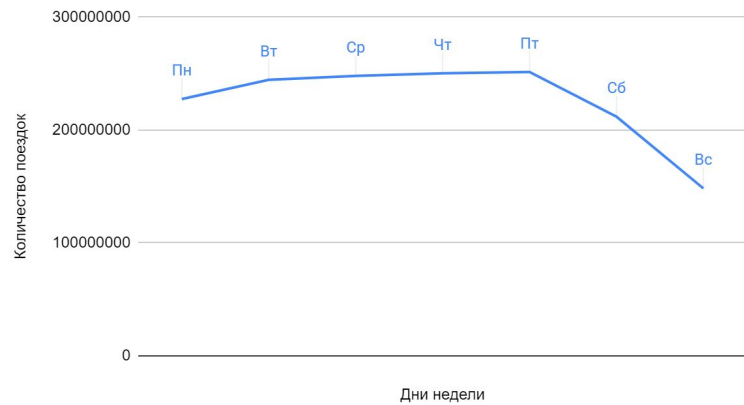




Автобусы

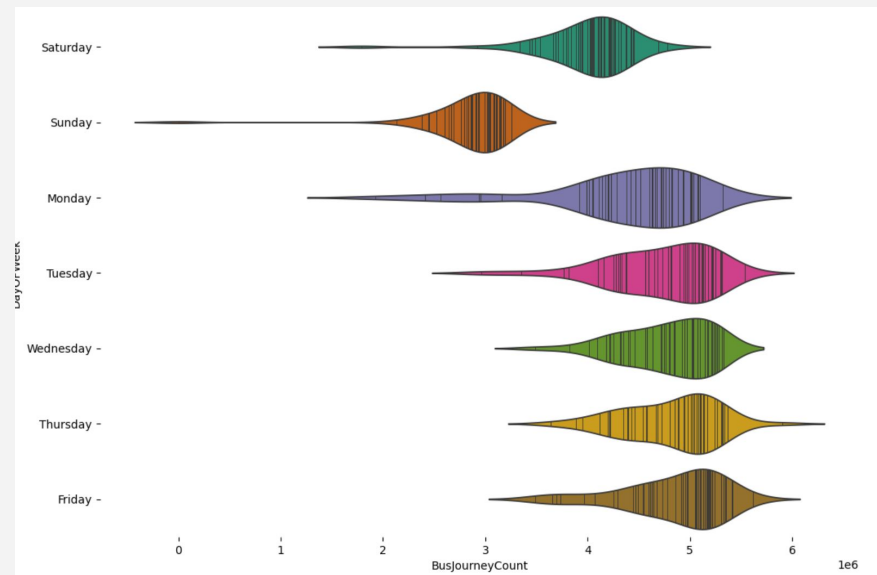
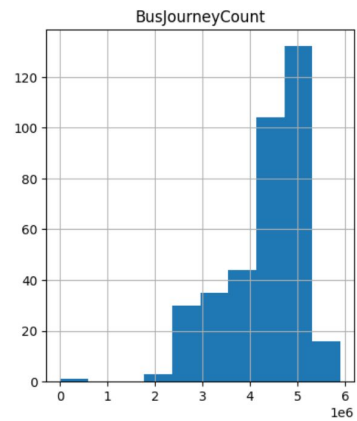
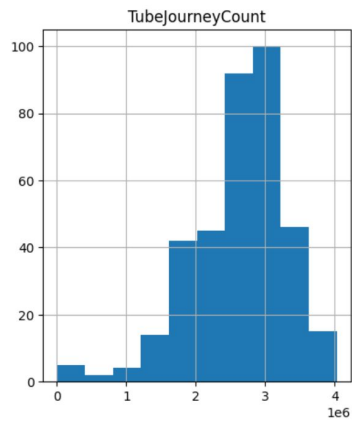


Автобусы



Распределение данных можно понять, если визуализировать их в виде гистограммы:

```
[ ] data.hist(data.columns, figsize=(10, 5));
```



Итоги

	Excel	Pandas	Итог
Визуализация данных	Широкий выбор инструментов	Визуальная составляющая может быть получена только с помощью сторонних библиотек	Excel
Скорость и производительность	Высокая скорость только для небольшого количества данных	Способность быстро выполнять задачи с большим количеством входных данных	Pandas
Сложность запросов и операций	Масштабные операции могут быть более сложными для выполнения, особенно при работе с данными, требующими сложных вычислений	Позволяет манипулировать большим объемом данных, в т.ч. несколькими базами одновременно, объединяя и разделяя их с помощью команд, не вручную	Pandas
Использование	Данные за счет форматирования могут быть наглядными, но тогда структура будет различная	Данные для баз всегда имеют фиксированный формат: столбцы и строки с заголовками и ID	Excel - для штучной работы Pandas - для задач машинного обучения

57

Человек были
заинтересованы проектом и
прошли опрос! Для них эта
проблема была актуальна!





Спасибо за внимание!

