

Индивидуальный проект

Цель проекта

На курсе «Визуализация данных и инфографика» основная задача студентов – научиться преобразовывать сырые данные в понятные и информативные визуализации с помощью чартов, используя подручные инструменты в виде современных программ и платформ для визуализации (Power BI, Tableau, Python и т.д.)

Данный проект направлен на:

- **Развитие навыков работы с данными:**
 - Поиск и оценка качества датасета
 - Очистка и преобразование данных с использованием Power Query в Power BI
- **Создание визуальных историй:**
 - Анализ данных и выявление ключевых инсайтов
 - Подбор адекватных типов визуализаций для каждого аспекта данных
- **Повышение уровня презентации данных:**
 - Обеспечение интерактивности и удобства восприятия данных
 - Формирование логичных и визуально привлекательных чартов для визуализации
- **Практическое применение теории в визуализации:**
 - Учет принципов теории цвета, композиции и восприятия данных
 - Обоснование выбора визуальных решений для представления информации

Целью данного проекта является обучение студентов комплексному подходу к процессу анализа данных, начиная с подготовки данных, заканчивая визуальным представлением в формате, подходящим для целевой аудитории

Этапы выполнения проекта

1. Поиск датасета:

- Найти доступный для использования датасет на свободную тему
- Рекомендуемые источники: Kaggle, Data.gov, Google Dataset Search, OpenData
- Датасет должен содержать не менее 1500 строк (после полной обработки данных) и 8-10 столбцов (с различным типами данных)

2. Очистка данных:

Использовать Power Query в Power BI для обработки данных:

- Удаление дубликатов
- Заполнение или удаление пропусков
- Приведение форматов данных (например, даты, числовые значения и т.д.)
- Создание новых столбцов (если необходимо для анализа)
- Применение фильтров для исключения нерелевантных данных

3. Анализ данных и визуализация:

- Создать **минимум 8 визуализаций**, которые могут раскрыть тему датасета
- Рекомендуемые типы графиков:
 - Столбчатые (Stacked column/bar chart, Clustered column/bar chart)
 - Линейные (Line chart, Area chart, Stacked area chart)
 - Круговые диаграммы (Pie chart, Donut chart)
 - Гистограммы
 - Карты (если данные содержат географические данные)

1) Каждая визуализация должна сопровождаться:

- Подписью, объясняющей суть данных
- Четко оформленными осями и легендой

Технические требования по проекту:

- **Язык проекта** – Русский, Английский
- **Инструменты** - Power BI (Power Query для очистки и визуализаций), Python по желанию
- **Форматы итоговых визуализаций:** Файл .pbix, отчет в формате PDF. Если работа выполнена на Python – ссылку и отчет.

Итоговый отчет

Для успешного выполнения проекта по структуре, описанной выше, необходимо создать отчет, в котором будет полное описание этапов вашей работы

Рекомендации по созданию отчета:

1. Титульный лист

- Название проекта
- ФИО студента
- Группа

2. Введение

- Кратко опишите тему вашего проекта и расскажите почему выбрали именно ее
- Опишите цель вашего проекта, т.е. какую задачу вы решаете с помощью анализа данных
- Пропишите задачи вашего проекта (например, 1) Очистка данных; 2) Создать визуализации для выявления ключевых инсайтов; и т.д.)

3. Описание датасета

- Расскажите из какого источника вы брали данные (например, Kaggle)
- Опишите общие характеристики вашего датасета:
 - общее количество строк и столбцов
 - описание ключевых переменных (например, «Столбец 'Age' содержит возраст участников). Можно представить переменные в виде таблицы
 - заострите внимание на том, какие данные могут быть полезны для анализа, а точнее, какие данные были использованы в ходе вашей работы

4. Очистка данных

Подробно опишите процесс обработки данных:

- Какие проблемы были выявлены в датасете (пропуски, дубликаты, несоответствие форматов и т.д.)

- Как эти проблемы были решены (например, «Пропуски в столбце 'Country' были заменены значением 'Not specified'»)
- Примеры преобразований (например, создание новых столбцов)

Также прикрепите скриншоты этапов очистки данных в Power Query

5. Анализ данных и визуализации

- Включите в отчет изображения созданных чартов
- Кратко объясните каждый график:
 - Какие данные представлены?
 - Какие выводы можно сделать?
 - Почему выбран именно этот тип визуализации?
- Старайтесь подчеркивать взаимосвязи между готовыми визуализациями

6. Выводы

- Опишите основные инсайты, полученные из анализа данных
- Ответы на вопросы, поставленные в начале, т.е. как данные помогают раскрыть тему
- Возможные рекомендации или сценарии использования результатов

7. Список использованных источников

- Если при создании отчеты были использованы доп. материалы из интернета – укажите ссылки на эти источники
- Если использовалась доп. литература или ресурсы, представленных мною на google drive – добавьте их в список и укажите страницы, которые помогли вам в работе

8. Приложения (при необходимости)

Включите дополнительные графики, таблицы или скриншоты, которые поддерживают анализ, но не вошли в основную часть отчета

Советы по созданию отчета

- Избегайте сложных формулировок текста, пишите четко и понятно
- Используйте единый шрифт и стиль при оформлении
- Не добавляйте лишней информации, которая не относится к проекту
- Перед сдачей отчета проверьте текст на орфографические и грамматические ошибки

Критерии оценивания:

2) Поиск и выбор датасета (10%)

- Соответствие датасета теме проекта: выбранные данные релевантны и позволяют выполнить поставленные задачи.
- Качество датасета: достаточно строк и столбцов для анализа (не менее 1500 строк и 8 столбцов)
- Описание датасета в отчете: указаны источник, структура, и основные переменные

3) Очистка данных (30%)

- Удалены дубликаты, заполнены или обработаны пропуски, устранены ошибки форматов.
- Логичные и обоснованные шаги очистки, такие как фильтрация или создание новых столбцов.
- Описание каждого этапа очистки с примерами или скриншотами из Power Query.

4) Визуализации данных (40%)

- Разнообразие графиков (20%): Использованы разные типы графиков, подходящие для представленных данных (не менее 8 визуализаций)
- Качество оформления (10%): Подписи, легенды, оси и заголовки присутствуют и понятны. Цветовая гамма и композиция соответствуют правилам визуализации данных
- Аналитическая ценность (10%): Каждый график раскрывает важный аспект данных. Выявлены ключевые тенденции, зависимости или аномалии.

5) Отчет (10%)

- Логичность структуры: отчет включает введение, описание датасета, процесс очистки, анализ визуализаций, выводы и список источников
- Грамотность: отсутствие ошибок в тексте, ясность и лаконичность изложения.
- Качество визуального сопровождения: графики вставлены четко, с подписями и пояснениями.

6) Выводы и аналитика (5%)

- Основные инсайты сформулированы четко и логично.
- Выводы связаны с целями проекта и подкреплены анализом данных.

7) Общий подход (5%)

- Творческий подход к выбору темы и выполнению задач
- Самостоятельность в выполнении работы (оригинальность)