**GeekBrains**

**Специальность: Аналитика**

**Факультет: Быстрый старт. Аналитик. Цифровые профессии.**

**ДИПЛОМНЫЙ ПРОЕКТ**

**«Анализ транспортной работы пассажирских авиакомпаний, осуществляющих вылеты из московских международных аэропортов Внуково, Домодедово, Шереметьево в период 01.01.2022 – 28.02.2023»**

Выполнил: Антропов К.С.

Москва 2023 г.

**Дипломный проект: «Анализ транспортной работы пассажирских авиакомпаний, осуществляющих вылеты из московских международных аэропортов Внуково, Домодедово, Шереметьево»**

**Содержание**

|  |  |
| --- | --- |
| Введение | 2 |
| Инструменты, используемые в работе | 4 |
| Описание кода в файле .ipynb | 7 |
| Описание работы по очистке данных | 10 |
| Описание элементов дашборда | 15 |
| Основные выводы по проведенному анализу | 20 |
| Заключение | 25 |
|  |  |

**Введение**

В современном развивающемся мире невозможно переоценить роль авиации, как единственной транспортной системы, связывающей между собой континенты  
и позволяющей в максимально короткие сроки добраться из пункта А в пункт Б (при расстоянии свыше 1000 км).

Авиация является драйвером развития экономики стран и регионов, обеспечивая регулярную перевозку людей и грузов в центры экономического роста, туристические регионы, а также прочие направления.

При этом, авиационная отрасль является наиболее уязвимой по причине высокой себестоимости полетов и влияния внешних глобальных факторов (эпидемии, глобальные катастрофы природного характера, внешне-политические факторы и т.п.)

На основании изложенного, мною принято решение о проведении анализа транспортной работы авиакомпаний для оценки работы авиакомпаний, осуществляющих вылеты из московских международных аэропортов, а также определить степень влияния внешних факторов на пассажирские авиаперевозки.

**Инструменты, используемые в работе**

Дипломная работа разделена на несколько основных этапов:

* анализ существующих сайтов и API для сбора информации   
  о выполненных рейсах;
* сбор данных о выполненных рейсах;
* очистка полученных данных;
* визуализация данных.

**Анализ существующих сайтов и API для сбора информации   
о выполненных рейсах.**

На данном этапе произведена оценка существующих решений по сбору данных о выполненных рейсах с помощью следующих инструментов:

1. Yandex Browser
2. Сеть Интернет

Были рассмотрены такие варианты как:

* Flightradar24 API (не подошел, т.к. на бесплатной версии существует ограничение по количеству предоставляемой информации, а также отсутствует возможность выгрузки данных за определенный период времени в разрезе аэропортов);
* сайты московских аэропортов dme.ru, svo.aero, vnukovo.ru (также не могли быть использованы в дипломной работе по причине отсутствия истории   
  о выполненных рейсах и разрозненности представляемой информации);
* сторонние сайты tablo-online.ru, rasp.yandex.ru (использование указанных сайтов также не представлялось возможным по причине отсутствия данных   
  о полетах, выполненных более месяца назад);
* сайт avia.mvsm.ru (данный сайт принят за основной сайт для сбора данных о выполненных полетах, т.к. имеется история полетов, начиная с 28 декабря 2021 года, содержимое сайта четко визуализировано и информация о рейсах сразу может быть собрана по всем московским аэропортам).

**Сбор данных о выполненных рейсах**

Сбор данных о выполненных рейсах с сайта avia.mvsm.ru осуществлен при помощи парсинга с применением следующих инструментов:

1. Jupyter Notebook, как основная среда разработки
2. Python, как основной язык программирования
3. Библиотеки requests, bs4, re – для парсинга данных с сайта
4. Библиотеки datetime, time – для работы со временем при формировании ссылок и установки времени ожидания при парсинге
5. Библиотека pandas – для создания датафрейма и предварительной очистки данных

**Очистка полученных данных**

Очистка данных в свою очередь также состоит из нескольких этапов:

1. Предварительная очистка данных (удаление дубликатов рейсов, очистка от совмещенных рейсов, приведение к единообразию типов воздушных судов, сортировка полученных данных и т.п.).

Основные инструменты: Python библиотека pandas

1. Итоговая очистка данных (приведение к единому формату названий аэропортов вылета и аэропортов прибытия, уточнение наименования авиакомпаний, кодов рейсов, дополнительное уточнение модели воздушного судна и т.п.).

Основные инструменты: Microsoft Excel с применением макросов

1. Подготовка данных для визуализации (уточнение форматов столбцов, расчет времени задержки рейса и т.п.).

Основные инструменты: Microsoft PowerBi.