

# **Анализ авиарейсов в США за 2015 год**

---

ноябрь 2022

Горохова Алевтина Александровна

# Оглавление

- О себе
- Описание проекта
- Анализ дальности полетов
- Распределение причин задержки рейсов по категориям и месяцам
- Распределение полетов по дням недели и месяцам
- Количество рейсов, средняя задержка вылетов по каждой авиакомпании
- Аэропорт с минимальной задержкой вылета
- Самая пунктуальная авиакомпания на прилет в Los Angeles International Airport
- Аэропорт, где самолёты проводят больше всего времени на рулении
- Модель отбора топ 3 аэропорта прилета (вероятность опоздания минимальная – RMSE метрика), в зависимости от аэропорта вылета
- Выводы и рекомендации по проекту

Nashville International Airport  
Austin-Bergstrom International Airport Detroit Metropolitan Airport  
Charlotte Douglas International Airport  
William P. Hobby Airport Fort Lauderdale-Hollywood International Airport Newark Liberty International Airport  
Seattle-Tacoma International Airport John Wayne Airport (Orange County Airport)  
Sacramento International Airport San Francisco International Airport LaGuardia Airport (Marine Air Terminal) Dallas Love Field  
Minneapolis-Saint Paul International Airport John F. Kennedy International Airport (New York International Airport)  
Phoenix Sky Harbor International Airport Dallas/Fort Worth International Airport  
Hartsfield-Jackson Atlanta International Airport  
Los Angeles International Airport Chicago O'Hare International Airport  
Miami International Airport Denver International Airport George Bush Intercontinental Airport  
Gen. Edward Lawrence Logan International Airport Orlando International Airport  
Philadelphia International Airport McCarran International Airport Ronald Reagan Washington National Airport  
San Diego International Airport (Lindbergh Field) Chicago Midway International Airport  
Baltimore-Washington International Airport Portland International Airport  
Salt Lake City International Airport Honolulu International Airport  
St. Louis International Airport at Lambert Field  
Oakland International Airport

# О себе

- Горохова Алевтина Александровна
- Контактная информация:
  - тел, WhatsApp **8(926)32-33006**
  - Telegram **@Alevtina\_Gorokhova**
- Образование: высшее, РЭА им Г.В. Плеханова факультет «Финансы и кредит»
- Руководитель ВСП 9040/00118 (г. Реутов) Восточного Головного отделения Среднерусского банка
- Основной функционал: наставничество сотрудников с целью выполнения плана офиса по нормативу производительности/приоритетным продуктам банка, обеспечение высокого качества сервиса. Постоянное отслеживание и прогнозирование бизнес-результата за счет анализа Дашбордов и выстраивания личной системы приоритетов по каждому продающему сотруднику
- Личные качества: командность, нацеленность на результат, принятие нового и внедрение в работу. По результатам 2021 года офис занял первое место в Чемпионате ИСУ по Отделению, как следствие оценка "А" за результативность по системе 5+.

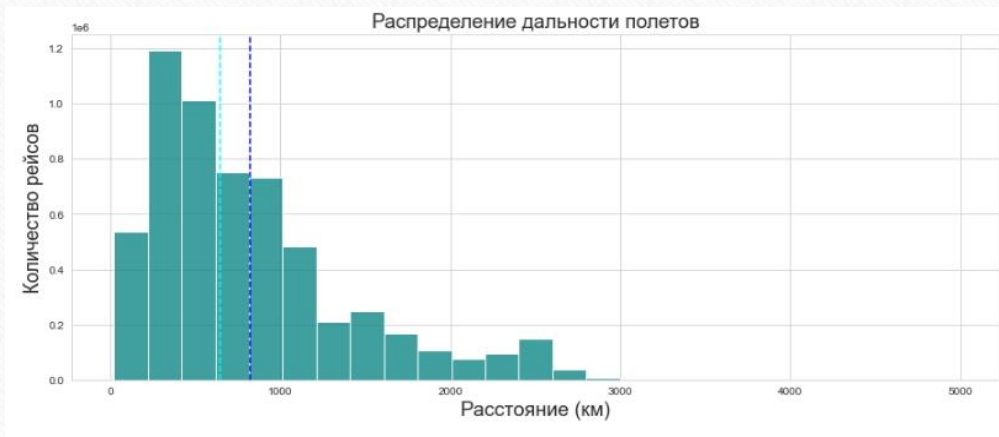


# Описание проекта

- Цель проекта – выявление факторов, влияющих на задержку рейсов, построение модели, которая будет выбирать топ 3 аэропорта прилета в зависимости от аэропорта вылета
- Ссылка на датасеты :  
<https://www.kaggle.com/datasets/usdot/flight-delays>
- Ссылка на github  
[https://github.com/gorokhovalevtina/DA/tree/main/Final\\_project](https://github.com/gorokhovalevtina/DA/tree/main/Final_project)
- Данные: основной датасет содержит сводную информацию о количестве своевременных, задержанных, отмененных и измененных рейсов в США за 2015 год (5.8 млн записей), два дополнительных файла с информацией об авиакомпаниях и аэропортах
- Реализованный процесс:
  - Подготовка, объединение данных
  - Группировка по признакам, расчет, визуализация
  - Построение модели
  - Выводы
- Используемые технологии:
  - Jupyter Notebook
  - Python (pandas, numpy, matplotlib, seaborn, sklearn, folium)

# Анализ дальности полетов

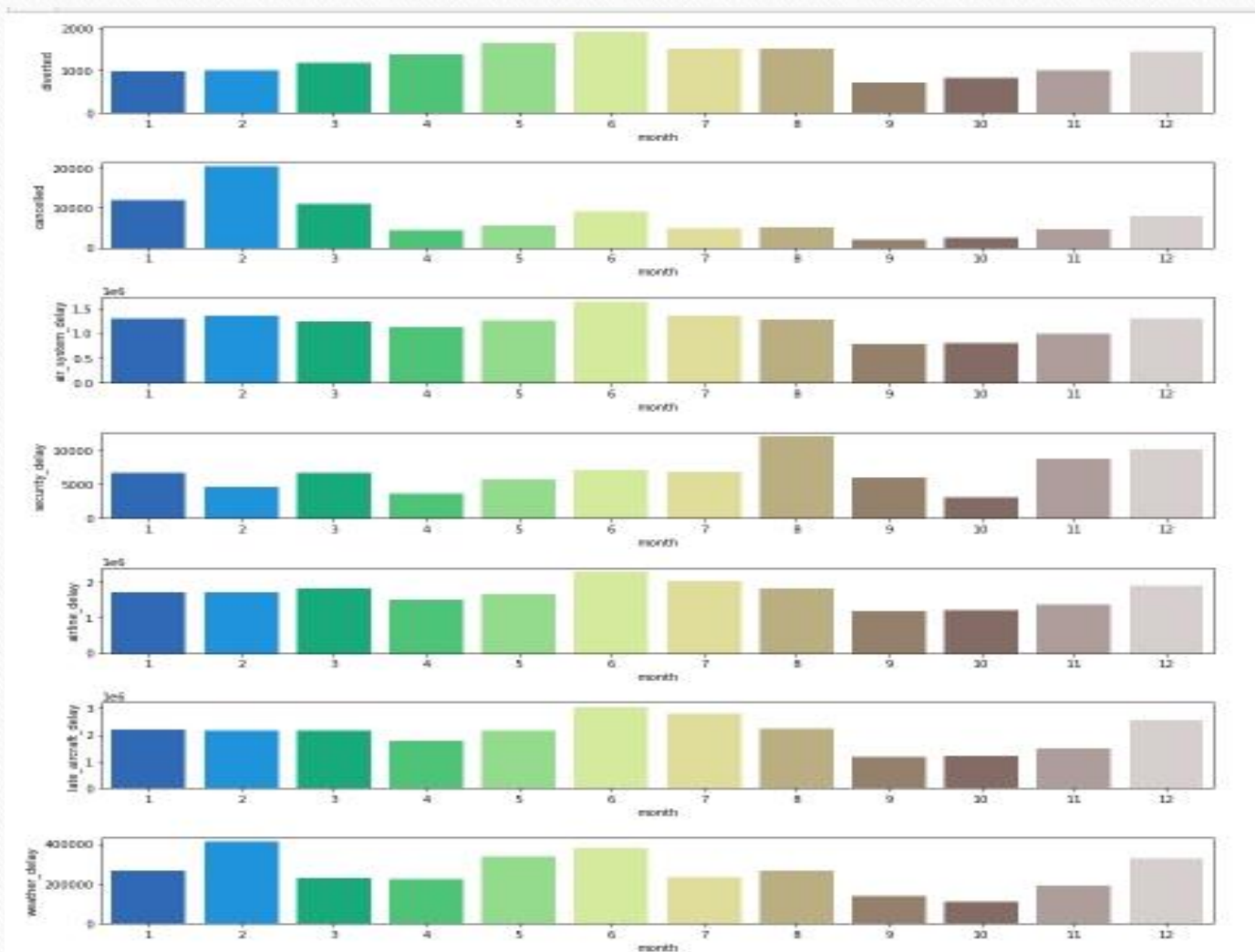
График представляет собой распределение с отклонением вправо. Наиболее часто полёты совершаются на расстояния от 200 до 800 км. На графике голубым цветом обозначена медиана, а синим - среднее значение дальности полетов. Они не равны, так как в среднее значение учитываются все достаточно длительные полёты, медиана же делит данные пополам.



Для данного графика данные были ранжированы по двум признакам: дальности полета(близко - до 700 км, далеко - свыше 700 км) и задержке прибытия (до 10 минут – минимальная задержка, более – рейс сильно задержан). В итоге получилось, что большинство рейсов относится к категории «Близкое расстояние\*Минимальная задержка»

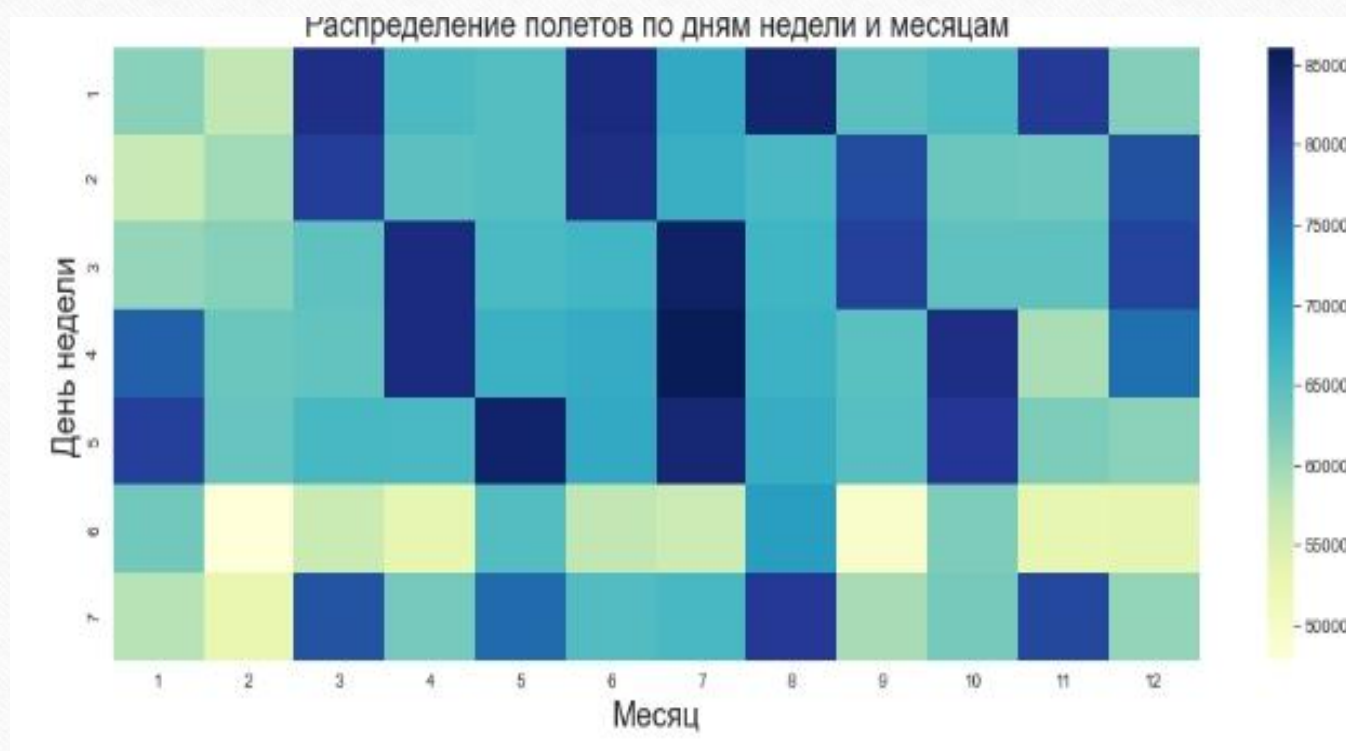
# Распределение причин задержки рейсов по категориям и месяцам

На графике можем видеть, что максимальное количество отмененных рейсов приходится на февраль, что в основном связано с погодными условиями. Также в декабре и феврале наблюдается большое число задержек из-за погоды, причина тому снежные заносы, а с сентября по ноябрь - наименьшее, поскольку погодные условия в этот период наиболее благоприятны. Что касается задержек в связи с безопасностью и перенаправления рейсов, то в период с мая по август наблюдаются максимальные показатели, что является следствием высокой загруженности аэропортов.





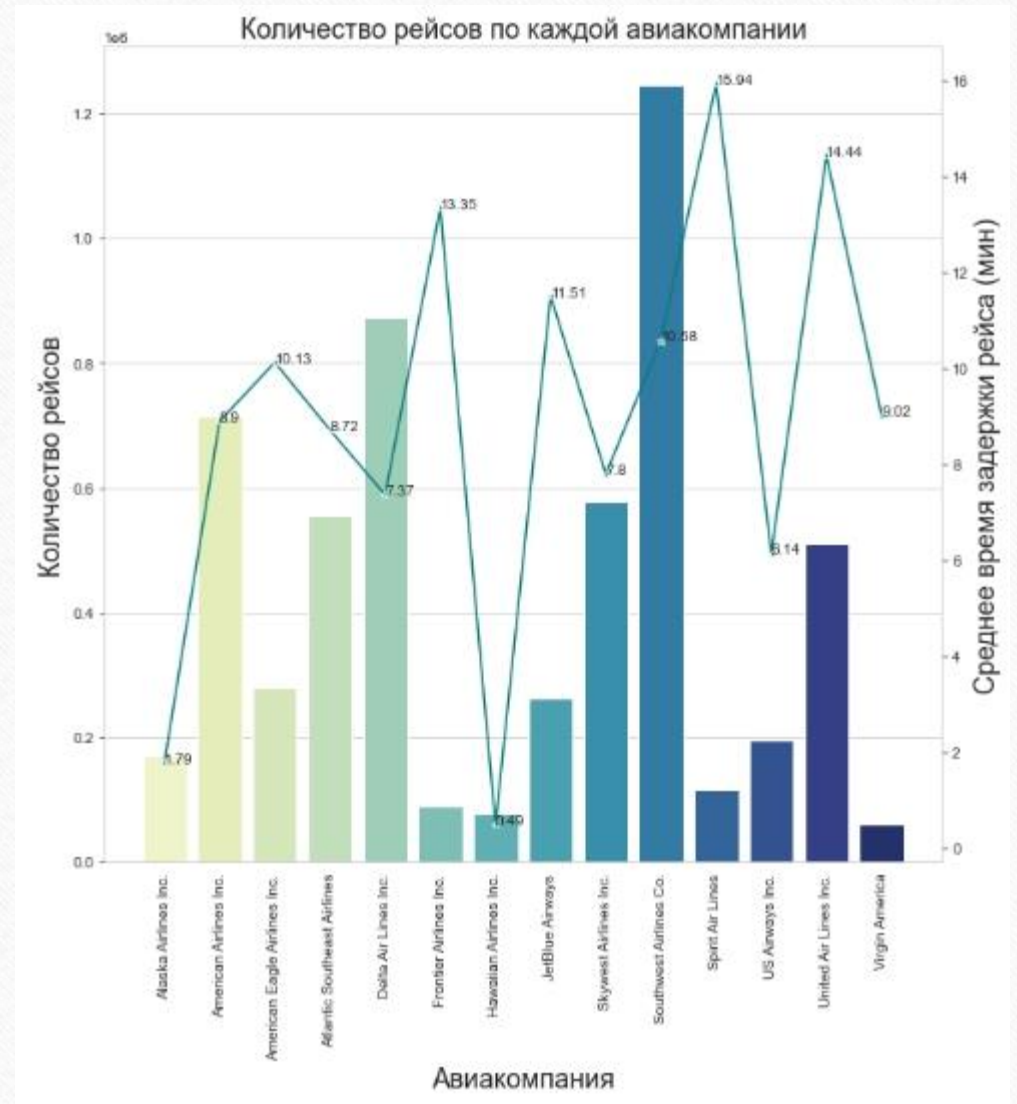
# Распределение полетов по дням недели и месяцам



Видно, что в субботу совершалось минимальное количество полетов, в то время как четверг является самым активным днем по числу перелетов. Причиной тому является большое количество бизнес-перелетов среди недели, в выходные же активность снижается. Если смотреть на месяца, то январь и февраль минимальны по данным показатели в связи с праздниками, а июль и декабрь максимальны, на них приходится пик летнего и зимнего отдыха.

# Количество рейсов, средняя задержка вылетов по каждой авиакомпании

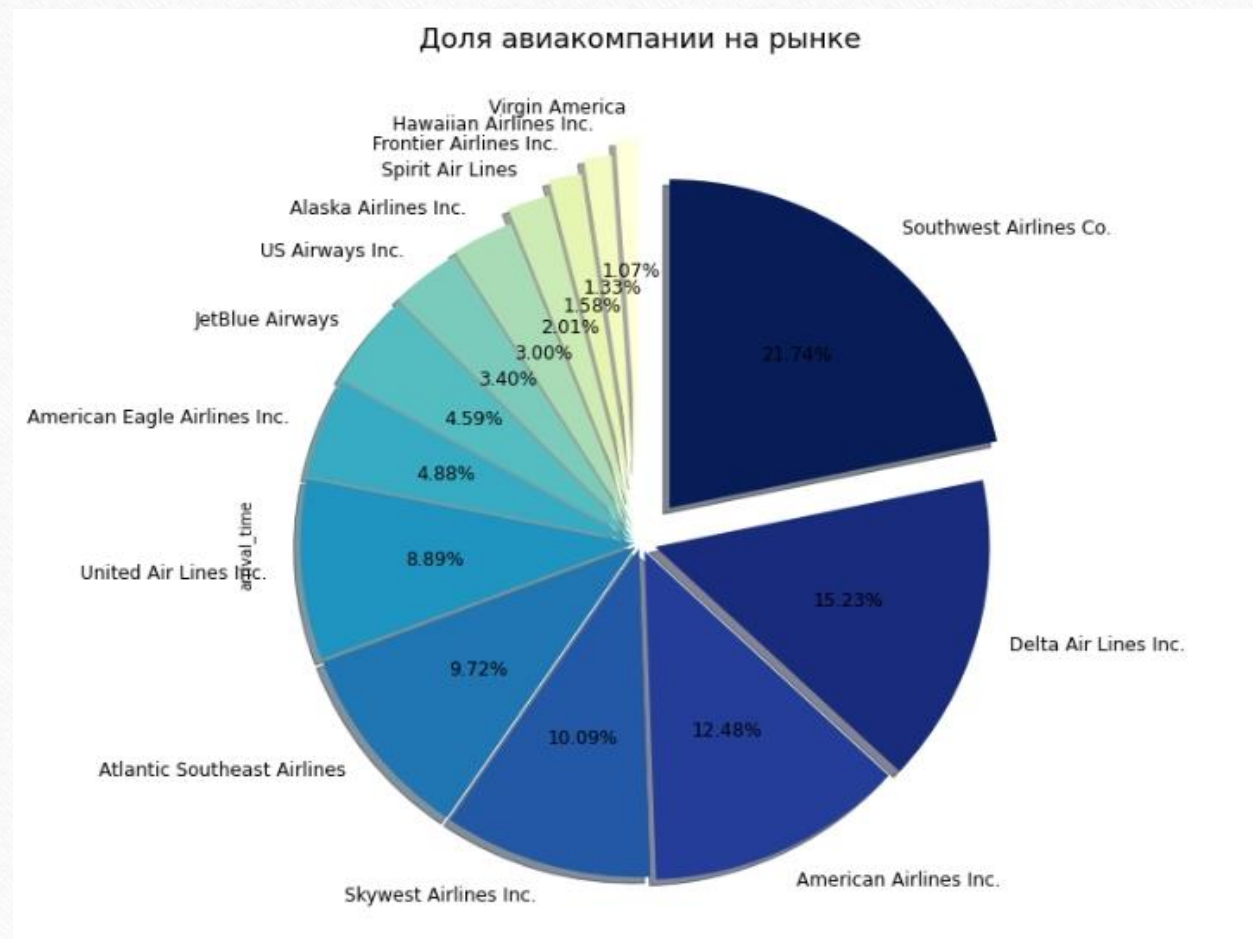
На графике показаны 14 авиакомпаний, а также количество предлагаемых ими рейсов и среднее время задержки. Количество рейсов не оказывает существенного влияния на задержки вылета, поскольку такие авиакомпании, как Spirit Air Lines, у которых не так много рейсов, имеют очень большое время задержки вылета. В то же время такие авиакомпании, как Southwest Airlines Co/Delta Air Lines Inc, которые предлагают большое количество рейсов, имеют гораздо меньшее время задержки вылета. Поэтому можно отметить, что Southwest Airlines Co/Delta Air Lines Inc - лучшие авиакомпании в сравнении с Spirit Air Lines.





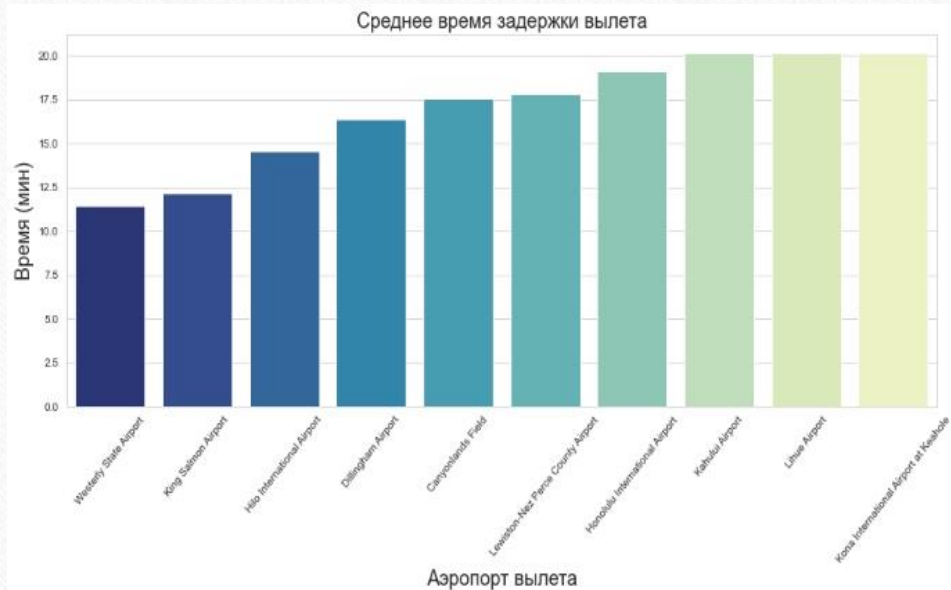
# Доля авиакомпаний на рынке

Половину на рынке авиаперевозок принадлежит Southwest Airlines Co. (21,74%), Delta Air Lines Inc(15,23%) и American Airlines Inc. (12,48%). Менее двух процентов у авиакомпаний Virgin America(1,07%), Hawaiian Airlines Inc. (1,33%) и Frontier Airlines Inc.(1,58%).

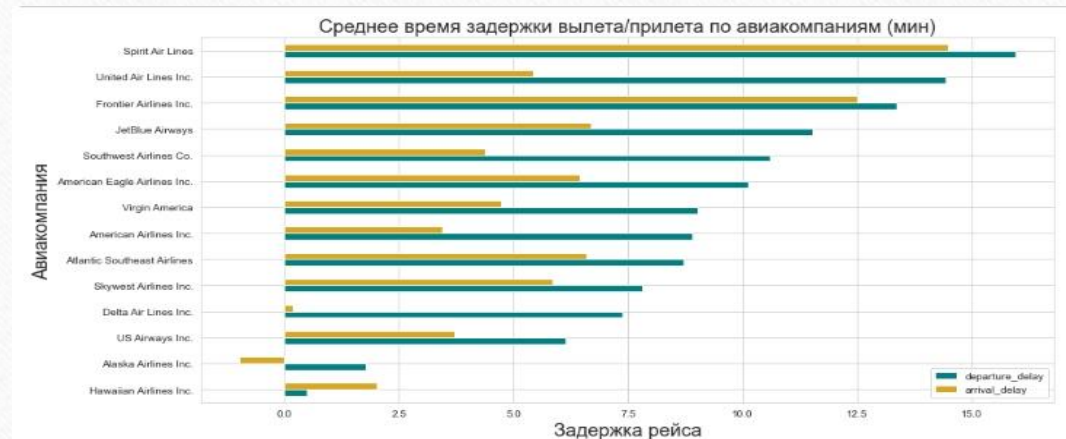


# Аэропорт с минимальной задержкой вылета

Westerly State Airport - аэропорт с минимальной задержкой вылета.

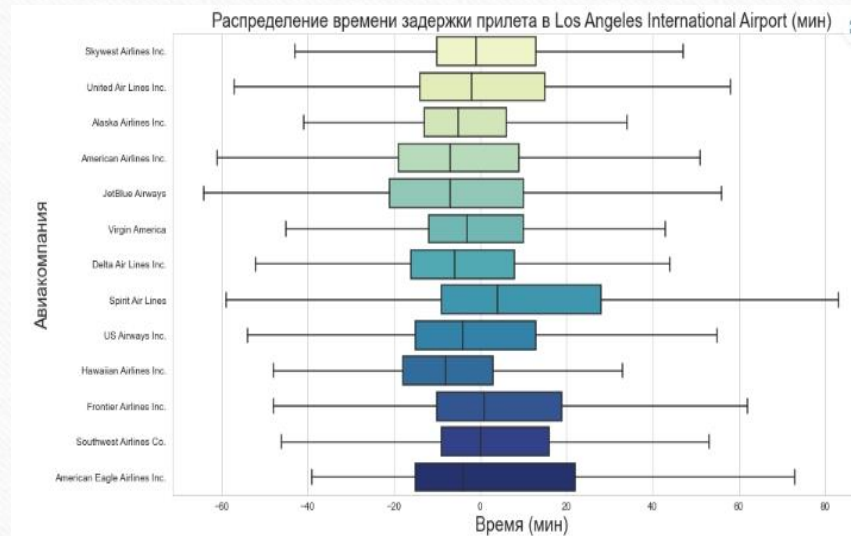
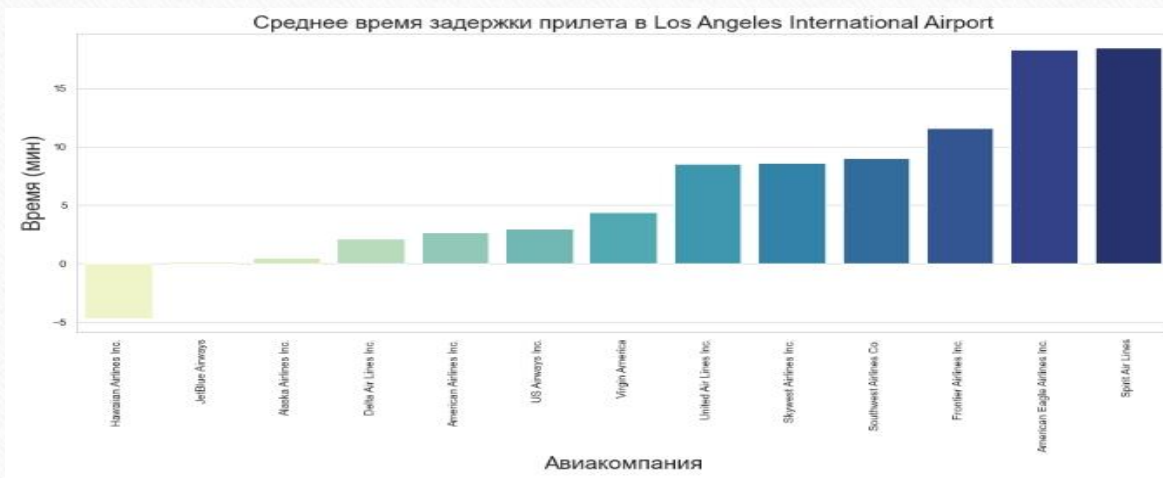


Время задержки вылета и прилета не всегда прямо коррелирует между собой. Особо выделяется Hawaiian Airlines Inc., где задержка вылета минимальна, а прилета достаточно высока и наоборот Delta Air Lines Inc, где задержка вылета по сравнению с задержкой прилета очень разнится.



# Самая пунктуальная авиакомпания на прилет в Los Angeles International Airport

Самой пунктуальной авиакомпанией с учетом среднего времени задержек и раннего прилета является JetBlue Airways, если учитывать пунктуальность только с точки зрения опоздания, то самой пунктуальной является Hawaiian Airlines Inc., если же учитывать пунктуальность только с точки зрения раннего прилета, то самой пунктуальной является Skywest Airlines Inc.

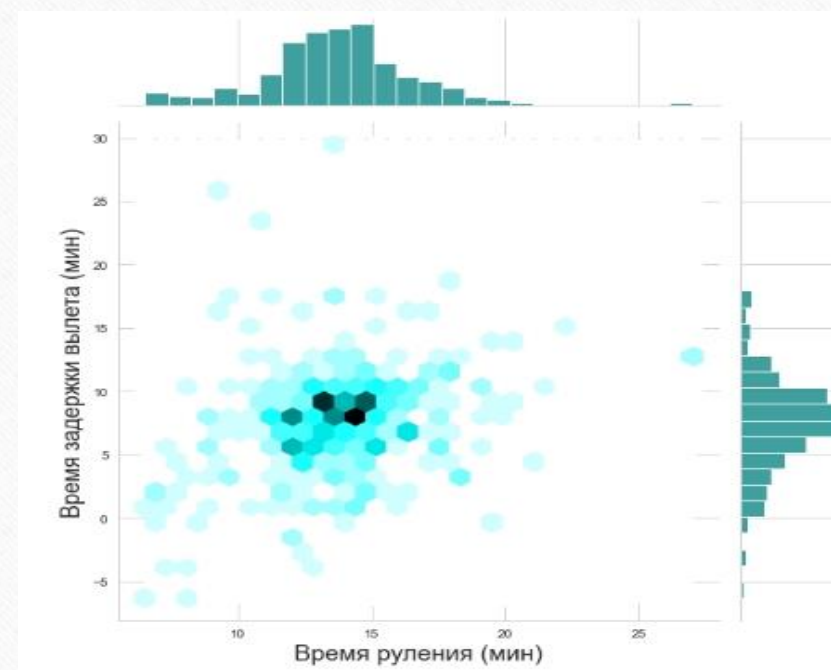
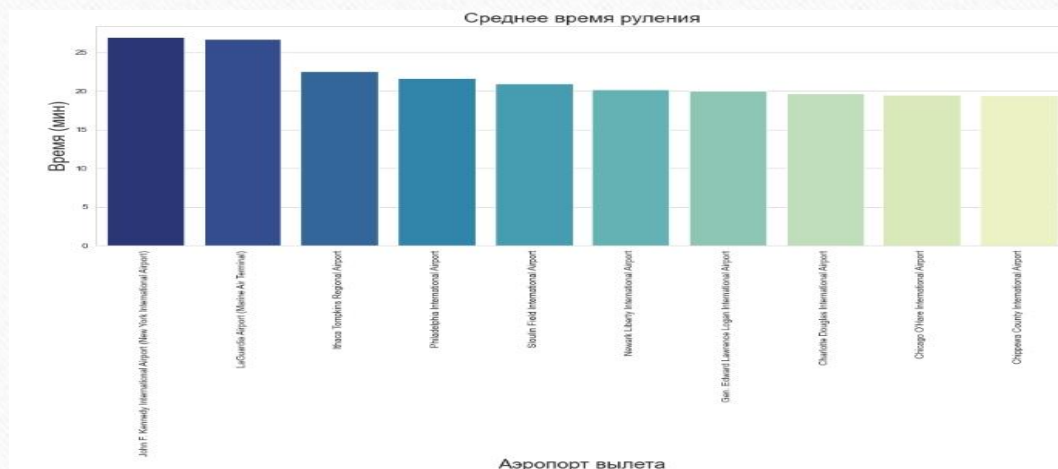


На графике выше также видно, что у Alaska Airlines Inc. минимальный разброс значений по задержке прилета, что говорит о стабильности ее работы. JetBlue Airways имеет минимальное среднее значение только за счет широкого диапазона значений.



# Аэропорт, где самолёты проводят больше всего времени на рулении

John F. Kennedy International Airport (New York) является аэропортом с максимальным средним временем руления перед взлетом. На графике справа не наблюдается прямой линейной зависимости времени руления и времени задержки вылета, значения сконцентрированы в диапазонах: время руления – от 12 до 16 минут, время задержки от 4 до 12 минут.



# Модель отбора топ 3 аэропорта прилета

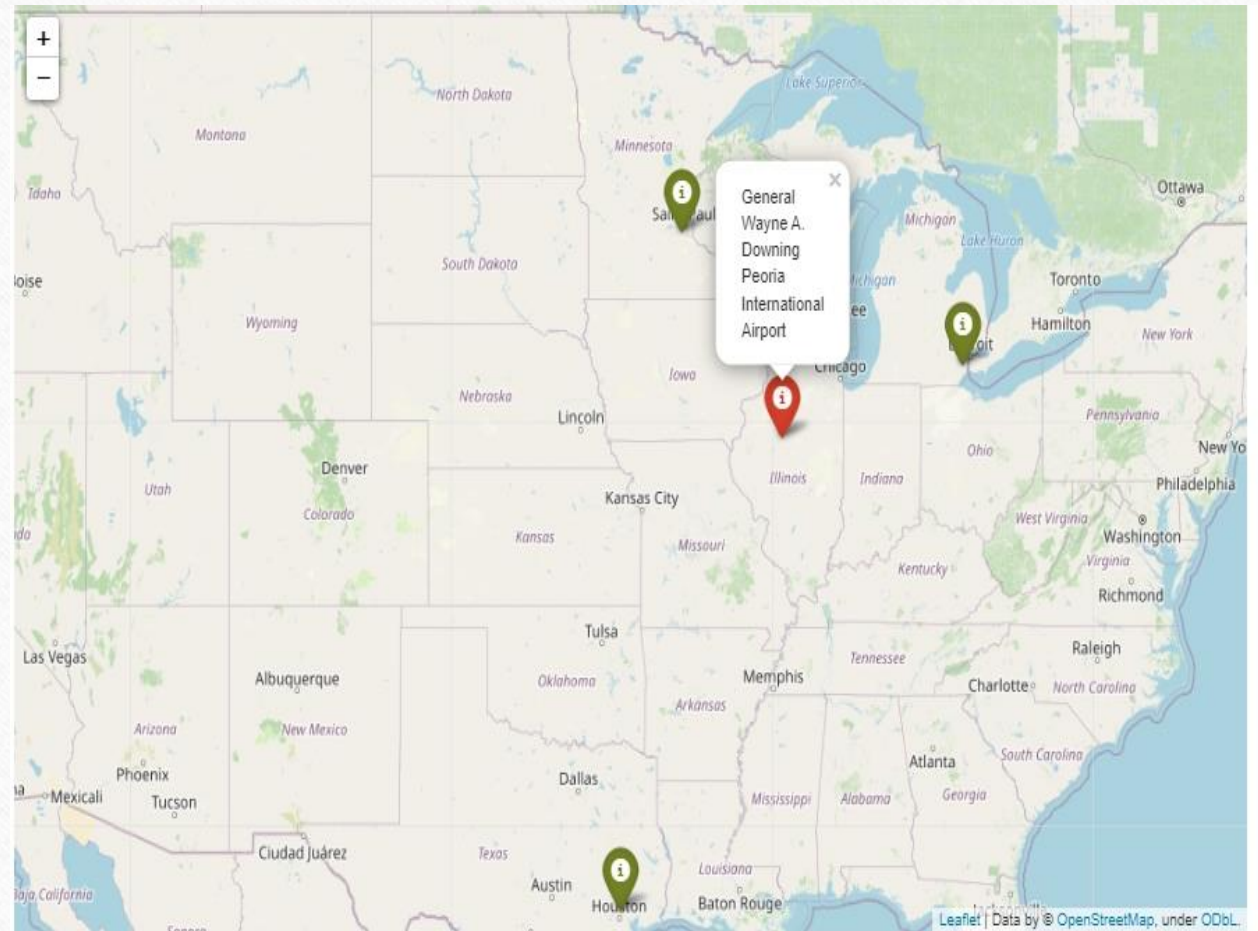
Лучшие направления с аэропорта General Wayne A. Downing Peoria International Airport

	destination_airport	mean_arrival_delay_in_past \
0	Detroit Metropolitan Airport	-11.142857
6	Minneapolis-Saint Paul International Airport	642.809524
5	George Bush Intercontinental Airport	1463.888889

RMSE

0	2339.192769
6	2923.240827
5	5199.863193

Пример работы модели по выбору топ 3 аэропорта прилета (вероятность опоздания минимальная – RMSE метрика), в зависимости от аэропорта вылета



# Выводы и рекомендации по проекту

- **Актуальность.** Ключевой целью любой авиакомпании является максимизация прибыли. Проведённый анализ может быть использован для составления рейтинга авиакомпаний, тем самым пользователи будут выбирать перевозчика с минимальными задержками, доход и доля компании на рынке будут расти. Вторым направлением использования является анализ внутренней организации работы авиакомпании и аэропорта, их взаимодействие, направленное на минимизацию задержек вылета/прилета. В-третьих, основные расчеты данного проекта могут быть актуальны и для анализа работы российских авиакомпаний и аэропортов.
- **Общие выводы.** Наиболее часто встречающийся диапазон дальности полетов составляет от 200 до 800 км, чаще всего самолеты прилетают вовремя либо с минимальными задержками на близкие расстояния. Задержки в основном связаны с погодными условиями. Southwest Airlines Co. (21,74%), Delta Air Lines Inc(15,23%) и American Airlines Inc. (12,48%) - топ 3 авиакомпаний, которые занимают половину рынка авиасообщения США. Количество рейсов и длительность руления не оказывает существенного влияния на задержки вылета. В субботу совершалось минимальное количество полетов, в то время как четверг является самым активным днем по числу перелетов. Январь и февраль минимальны по данным показатели в связи с праздниками, а июль и декабрь максимальны, на них приходится пик летнего и зимнего отдыха.



Спасибо за внимание!