**IOI Project 1st**  
**January Flight Delay Prediction**

**Day : 2020. 06. 08**

**Name : 김승주**

* **Dataset :** [**https://www.kaggle.com/divyansh22/flight-delay-prediction**](https://www.kaggle.com/divyansh22/flight-delay-prediction)
* **설명**

This is the first part of flight delay prediction i.e. for the month of January. To check the second part i.e. for the month of February, please have a look at this dataset: [February Flight Delay Prediction](https://www.kaggle.com/divyansh22/february-flight-delay-prediction)

- 이것은 1월 비행지연예측의 첫번째 부분(즉, 한달동안)입니다. 2월 두번째 부분(즉, 한달)을 확인하고 싶다면 이 데이터셋을 통해 확인하세요

\*i.e. : in other words , 다시말하면, 즉

This data is collected from the Bureau of Transportation Statistics, Govt. of the USA. This data is open-sourced under U.S. Govt. Works. This dataset contains all the flights in the month of January 2019 and January 2020. There are more than 400,000 flights in the month of January itself throughout the United States. The features were manually chosen to do a primary time series analysis. There are several other features available on their website.

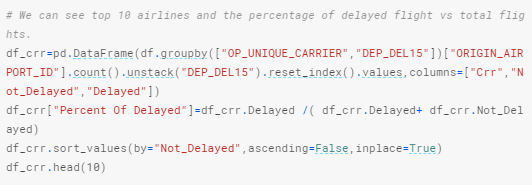
* 이 데이터는 미국 교통통계국(Bureau of Transportation Statistics, Govt. of the USA)를 통해 수집되었습니다. 또한 미국정부 산하에 오픈소스로 공개되어 있습니다. 이 데이터집합은 2019년 1월, 2020년 1월의 모든 항공편을 포함하고 있습니다. 그곳엔 미국의 모든 항공편을 합쳐서 1월 동안 비행했던 40만 항공편 이상 데이터를 포함하고 있습니다. Primary time series analysis의 기능을 하기 위해 선택하게 되었습니다. 그 웹사이트에서는 몇가지 다른 기능을 제공합니다.

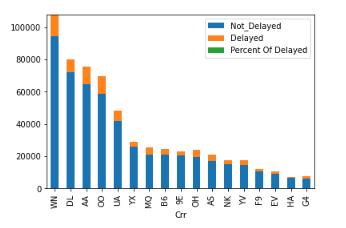
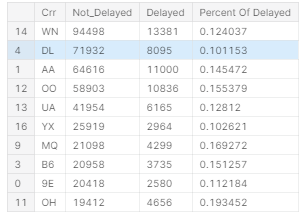
This data could well be used to predict the flight delay at the destination airport specifically for the month of January in upcoming years as the data is for January only.

* 이 데이터는 이번해의 1월동안에만 있었던 데이터를 통해서 1월에 일어날 목적지에서의 항공편 지연을 미리 예방할 수 있다
* **데이터 칼럼**
* **DAY\_OF\_MONTH :** 한달 단위의 하루 항공지연 빈도
* **DAY\_OF\_WEEK :** 일주일 단위의 하루 빈도
* **OP\_UNIQUE\_CARRIER :** 항공사 코드
* **OP\_CARRIER\_AIRLINE\_ID :** 항공사의 고유 ID 코드
* **OP\_CARRIER :** 항공권 예약과 스케줄에 따른 IATA(국제항공운송협회)에 의한 가변적인 항공사 코드
* **TAILNUM :** 항공기 등록 번호
* **OP\_CARRIER\_FL\_NUM :** 항공편 번호
* **ORIGIN\_AIRPORT\_ID :** 미국 교통부에서 지정한 고유 식별 번호
* **ORIGIN\_AIRPORT\_SEQ\_ID :** 미국 교통부에서 특정 시점일때 지정한 고유 식별 번호
* **ORIGIN :** 출발지 공항
* **데이터 이해하기**
* 데이터 입력

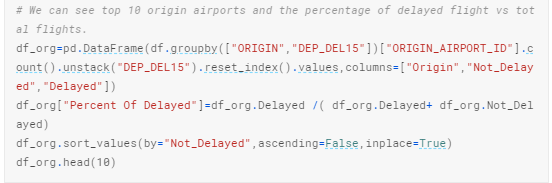


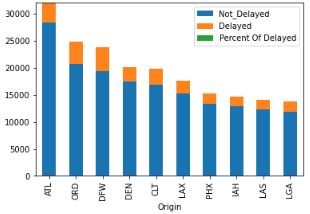
* 항공사에 따른 정상 항공편과 지연된 항공편



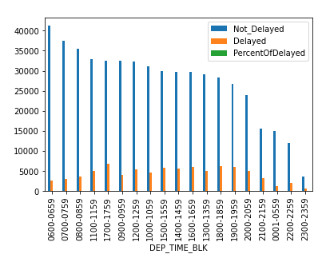


* 공항에 따른 정상 항공편과 지연된 항공편



* 시간 순서에 따른 지연



* **데이터 결론**
* 항공사의 수요가 많을수록 항공편의 지연이 생기는 일도 증가함
* 공항의 수요와 항공편의 지연이 발생 빈도는 상관관계가 작지만 크게 두 부분으로 나누어 보면 수요가 작은 공항에서는 일정한 지연항공편이 생기지만, 일정 수요 이상인 공항에서는 공항에 따른 지연률이 다르다는 것을 알 수 있다.
* 시간순서에 따른 지연빈도를 통해 어떤 날에는 월등히 많고 적은 것을 통해서 공항과 항공편에 따른 것이 아닌 제 3의 외부요인으로 항공편의 지연이 일어난다는 것을 알 수 있음.(ex. 공항 별 날씨)
* **데이터 활용 방안 제시**
* 공항의 날씨 데이터와의 분석을 통해 공항 별 날씨를 통해 일어나는 항공편 지연 파악하고 날씨에 따른 지연 예상시간 예측하여 알림
* 수요가 많은 항공사와 지연이 많이 발생하는 공항과의 스케줄 공유로 미리 지연시간을 예측하고 이용자에게 알림