

# HIVE 및 SQL 실습

## Airdelay table 사용

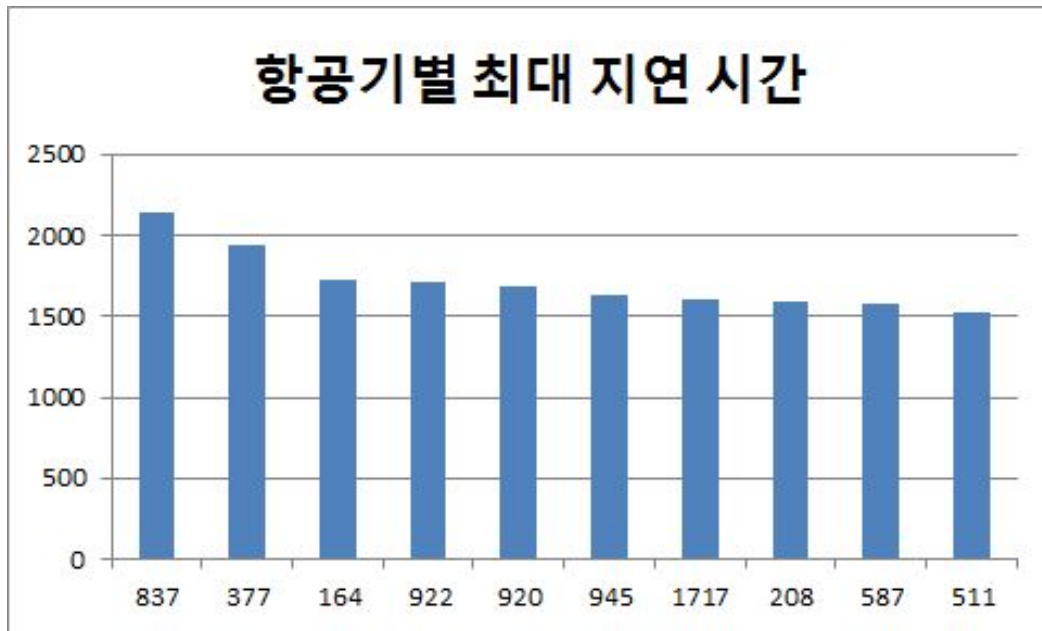
박수민

# 1. 최대 지연을 일으킨 비행기를 찾아라.

```
> select flightnum, max(arrdelay) maxd  
> from airdelay  
> group by flightnum order by maxd desc  
> limit 10;
```

**1999~2002년 사이에**

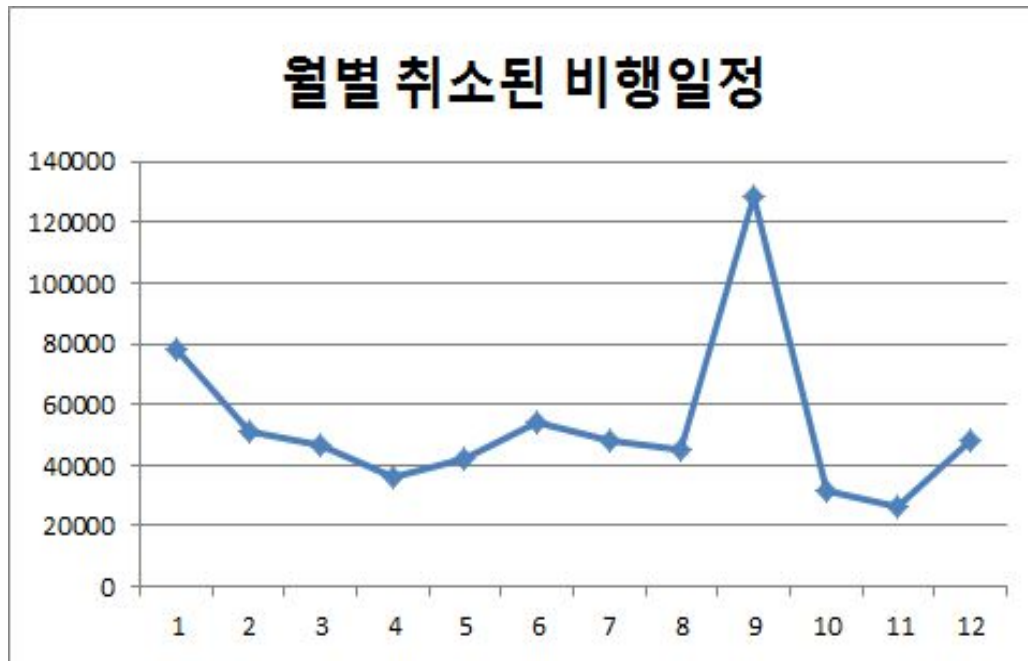
**항공기 837이 최대 지연시간을 보인다.**



## 2. 최다취소된 비행일정은 몇 월에 주로 나타나는가?

- > select month, sum(cancelled) as can
- > from airdelay
- > group by month order by can;

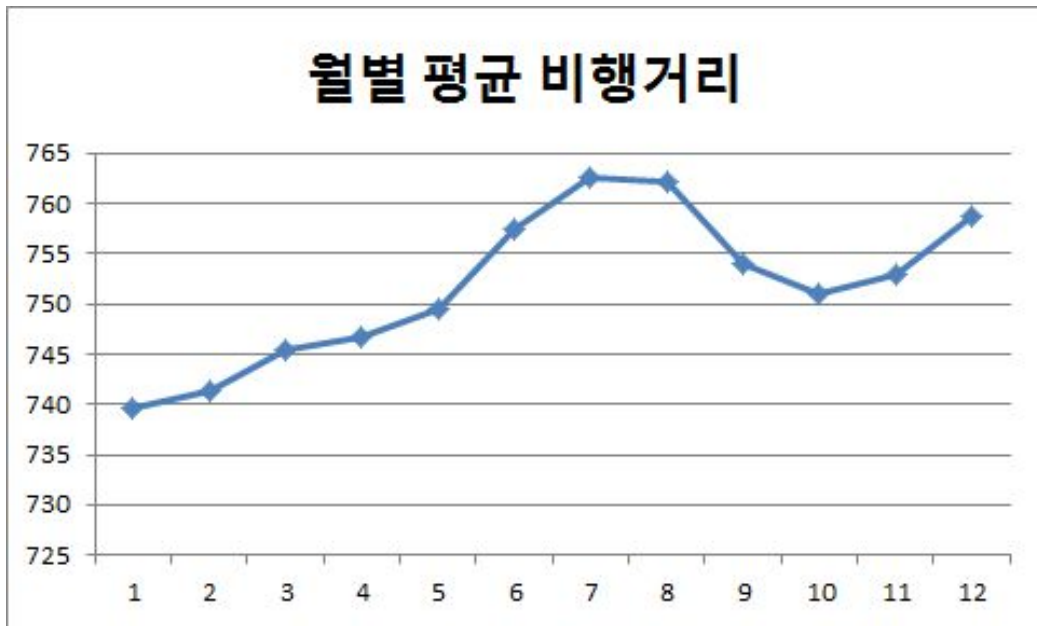
**1999~2002년 사이 9월에 비행  
취소가 가장 잦은 것으로 나타난다.**



## 4. 월별 비행거리가 차이가 있는지 살펴보시오?

- > select month, avg(distance) as avgD
- > from airdelay
- > group by month order by avgD desc;

**1999~2002년 사이 7월에  
평균 비행거리가 가장 긴 편이다.**



## 5. 년도별 취소 비행건수를 비교하시오.

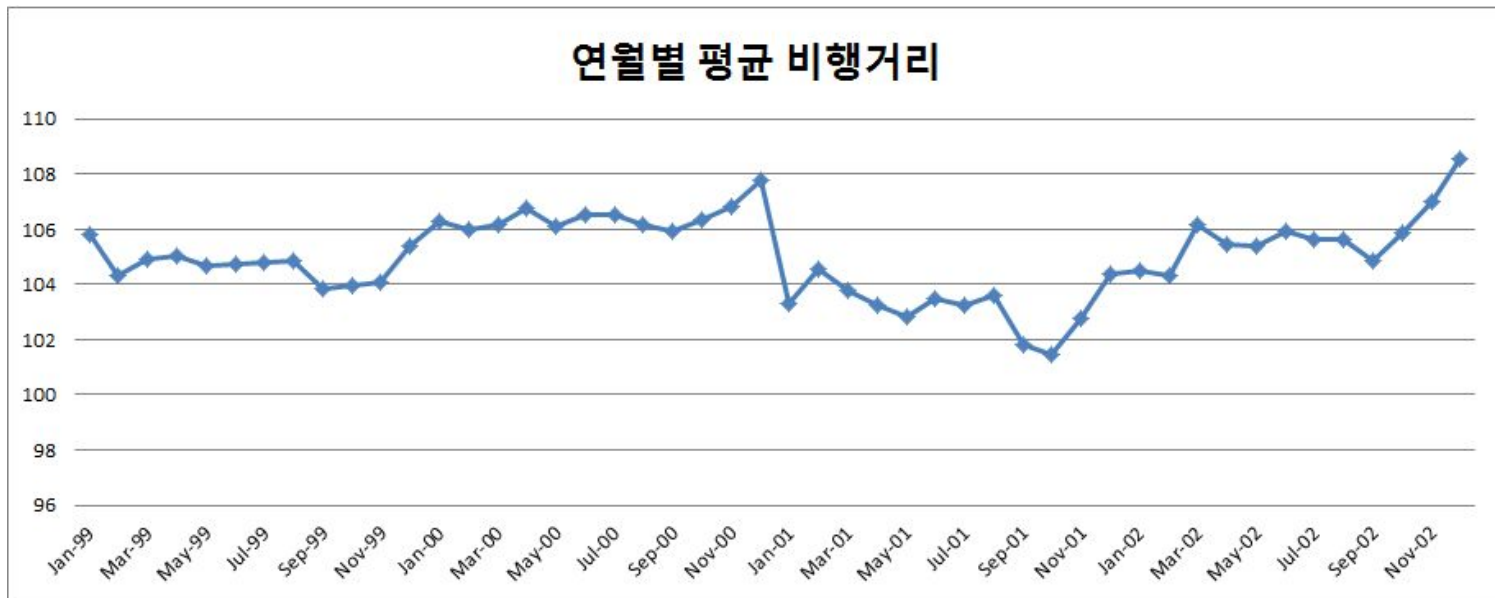
- > select year, sum(cancelled) as can
- > from airdelay
- > group by year order by year;



**1999~2002년 사이 2001년에  
비행취소가 가장 잦은 것으로 나타난다.**

## 6. 평균비행시간을 년, 월별로 나타내시오.

- > select year, month, avg(airtime) as avg
- > from airdelay
- > group by year, month order by year, month;

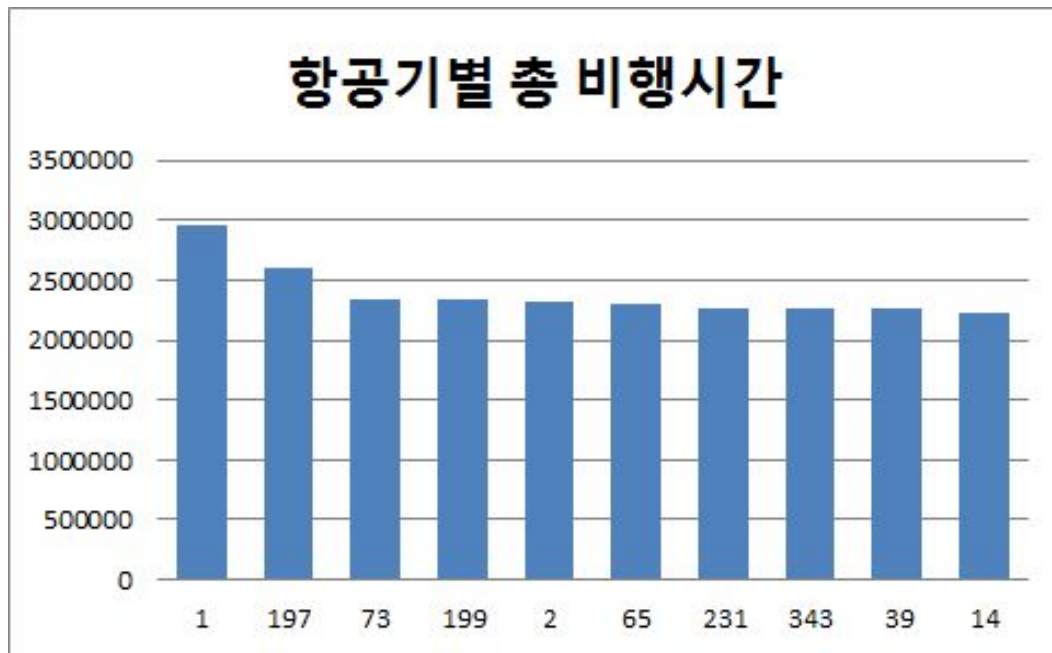


## 7. 항공기별 비행시간을 오름차순 나열. (10개 한정)

```
> select flightnum, sum(airtime) as atime  
> from airdelay  
> group by flightnum order by atime desc  
> limit 10;
```

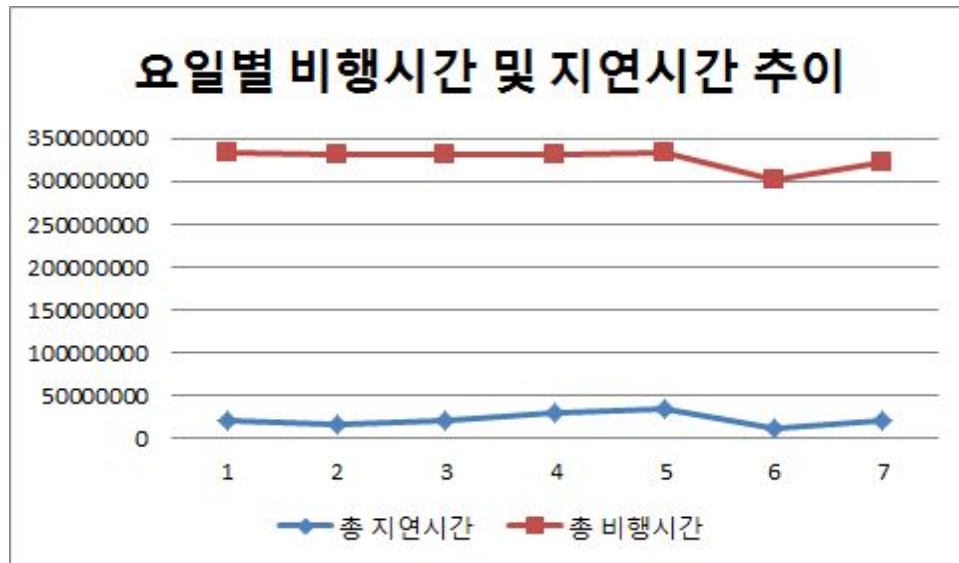
**1999~2002년 사이**

**항공기 1번이 가장 높은 총 비행시간을 보유한다**



## 8. 요일별로 지연시간과 비행시간을 비교하라.

```
> select dayofweek, sum(arrdelay) as d, sum(airtime) as a  
> from airdelay  
> group by dayofweek;
```



**1999~2002년 사이**

주중 5번째 요일에 비행시간과 지연시간 모두가 최댓값을 가진다.

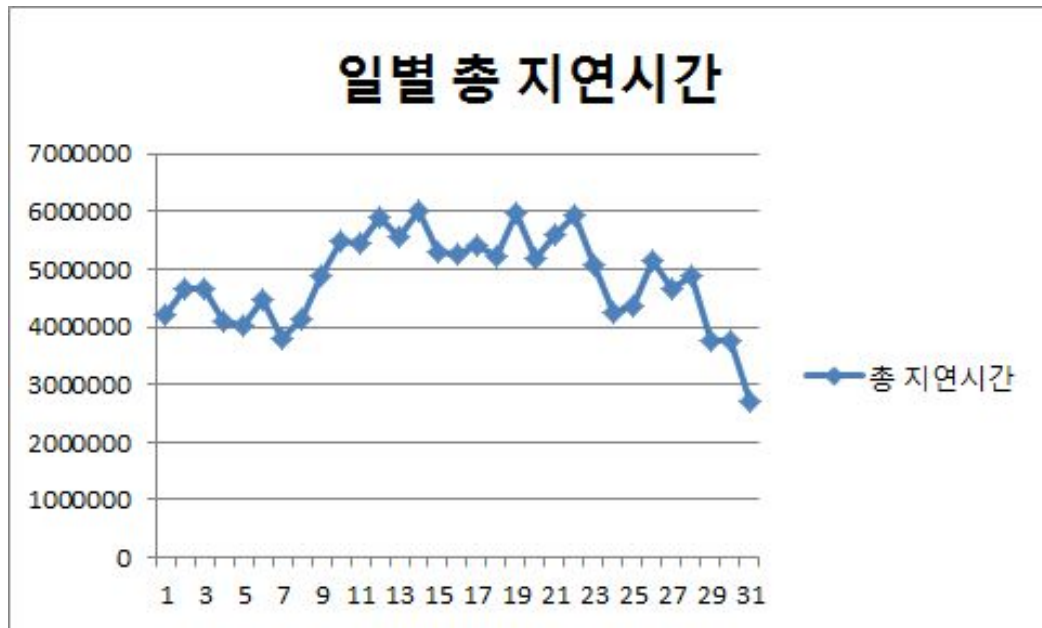


## 9. 매월중 일별로 지연시간을 비교하라

```
> select dayofmonth, sum(arrdelay) as delay  
> from airdelay  
> group by dayofMonth order by dayofmonth;
```

**1999~2002년 사이**

**월별 대략 11-15일 사이와 18-20일 정도에 가장 지연 시간이 높다.**

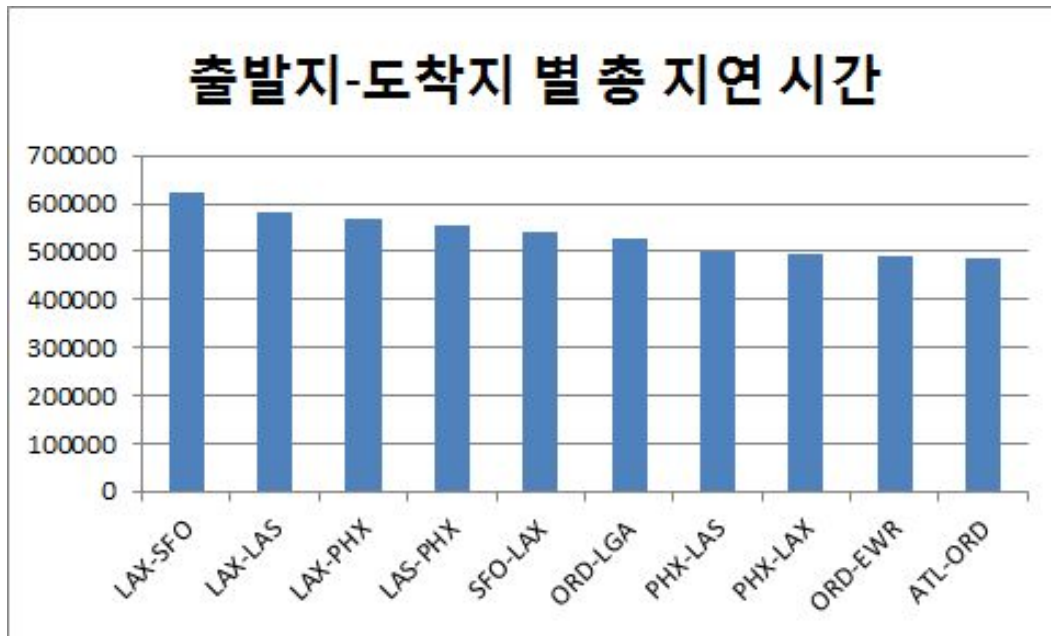


## 10. 출발지-도착지별 지연시간 총합 비교 [10개 추출]

```
> select origin, dest, sum(arrdelay) as ad  
> from airdelay  
> group by origin, dest order by ad desc  
> limit 10;
```

**1999~2002년 사이**

**LAX에서 출발하여 SFO로 향하는 비행이 가장 지연이 길었다.**

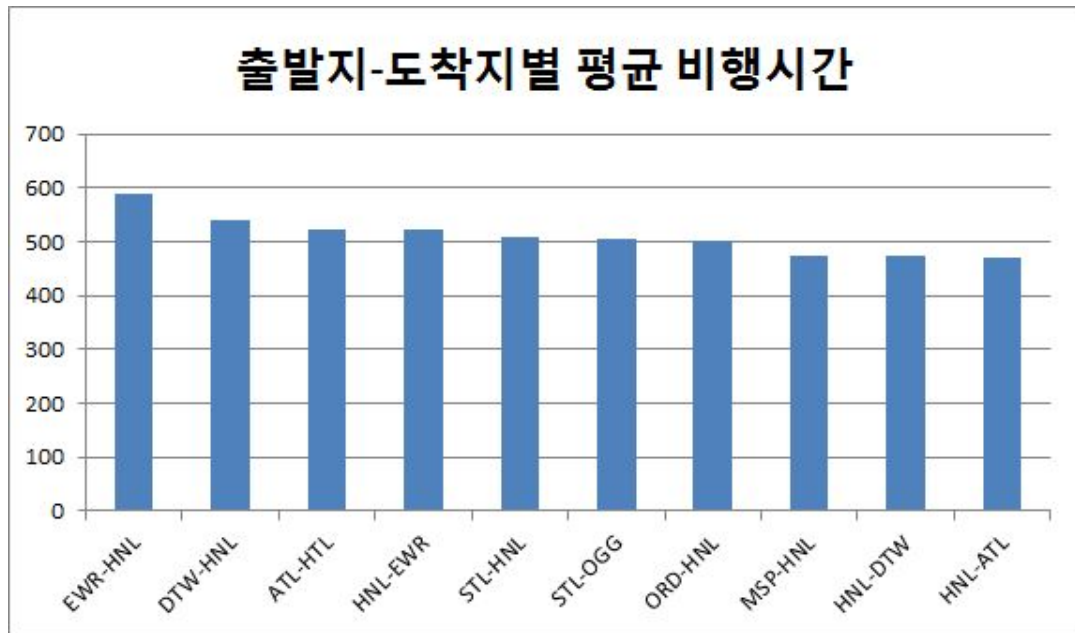


## 11. 출발지-도착지 평균 비행시간 내림차순 [10개 추출]

```
> select origin, dest, avg(airtime) as avg  
> from airdelay  
> group by origin, dest order by avg desc  
> limit 10;
```

**1999~2002년 사이**

**EWR에서 출발하여 NHL로 향하는 항공의 평균 비행시간이 가장 길었다.**



## 12. 연도별 요일별로 비행시간 차이가 있는가.

> select year, dayofweek, avg(airtime) as avg from airdelay

> group by year, dayofweek order by year, dayofweek;

연도별 요일별 평균 비행시간

