

HIVE 예제문제

Lotte Table

김영진

테이블 생성 및 데이터 읽기

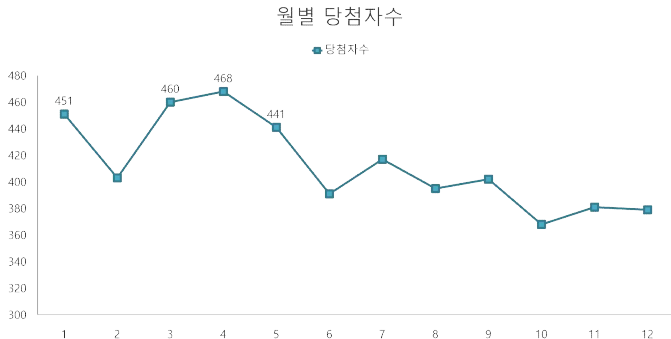
테이블 생성하기

```
create table lotto (  
  seq_no string,  
  win_no1 int,  
  win_no2 int,  
  win_no3 int,  
  win_no4 int,  
  win_no5 int,  
  win_no6 int,  
  win_no7 int,  
  winner_nu int,  
  year int,  
  month int)  
row format delimited  
fields terminated by ','  
lines terminated by '\n'  
stored as textfile;
```

데이터 불러오기

```
load data local inpath 'lotto.csv'  
overwrite into table lotto;
```

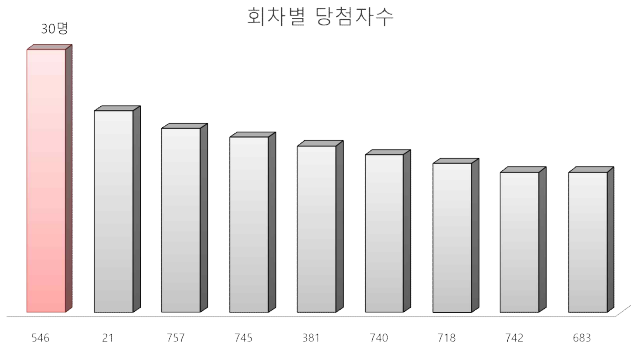
Q1. 월별 당첨자수를 구분하여 나타내시오



```
select month, winner_nu  
from lotto  
group by month;
```

6월 이전에 비교적 많은 당첨자수가 발생

Q2. 회차별로 당첨자수를 내림차순으로 나타내시오

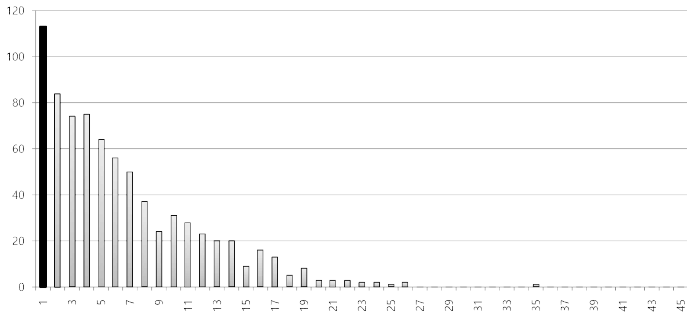


```
select seq_no, winner_nu  
from lotto  
order by winner_nu desc;
```

546회차에 당첨자수가 가장 많음

Q3. 각 번호의 선택숫자에서 나타나는 숫자들의 패턴을 파악하시오

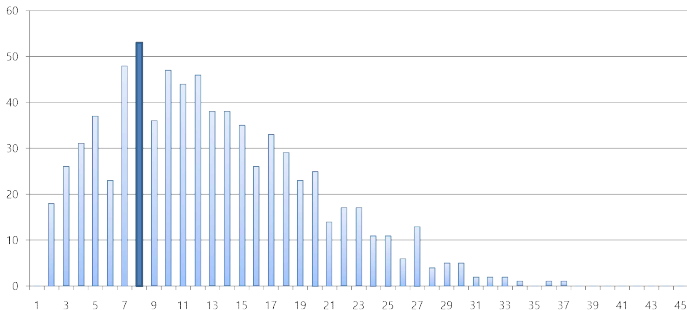
첫번째 로또번호 분포



첫번째 로또번호는
주로 낮은 숫자(1~7)의 빈도가 많은 경향을 보임

Q3. 각 번호의 선택숫자에서 나타나는 숫자들의 패턴을 파악하시오

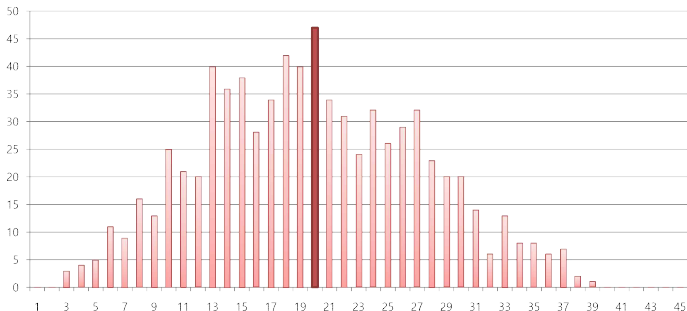
두번째 로또번호 분포



두번째 로또번호는
7,8,10,11,12,13번호가 비교적 많다

Q3. 각 번호의 선택숫자에서 나타나는 숫자들의 패턴을 파악하시오

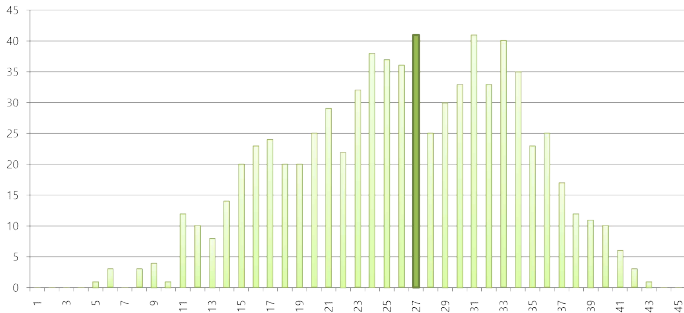
세번째 로또번호 분포



세번째 로또번호는
20번이 가장 많았으며,
13, 14, 15, 18, 가 비교적 많다

Q3. 각 번호의 선택숫자에서 나타나는 숫자들의 패턴을 파악하시오

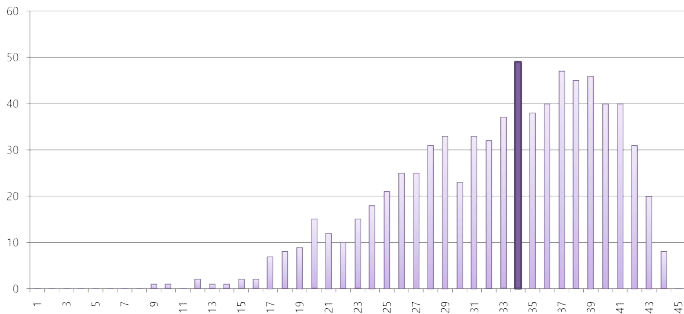
네번째 로또번호 분포



네번째 로또번호는
27번이 가장 많았으며,
31,33 번호가 비교적 많다

Q3. 각 번호의 선택숫자에서 나타나는 숫자들의 패턴을 파악하시오

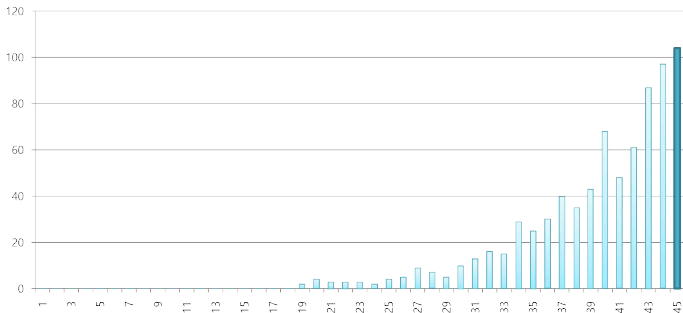
다섯번째 로또번호 분포



**다섯번째 로또번호는
34번이 가장 많았으며, 37~39번호가 비교적 많다**

Q3. 각 번호의 선택숫자에서 나타나는 숫자들의 패턴을 파악하시오

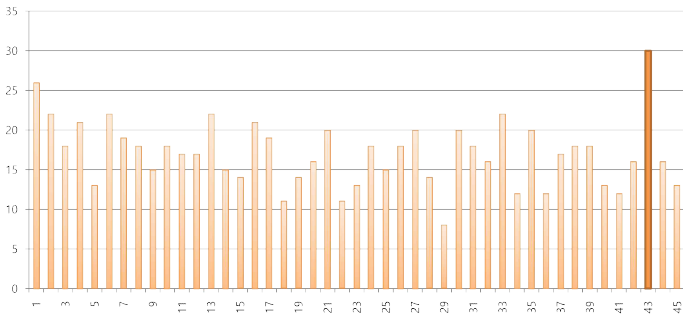
여섯번째 로또번호 분포



여섯번째 로또번호는
45번이 가장 많았으며, 40번대 번호가 비교적 많다

Q3. 각 번호의 선택숫자에서 나타나는 숫자들의 패턴을 파악하시오

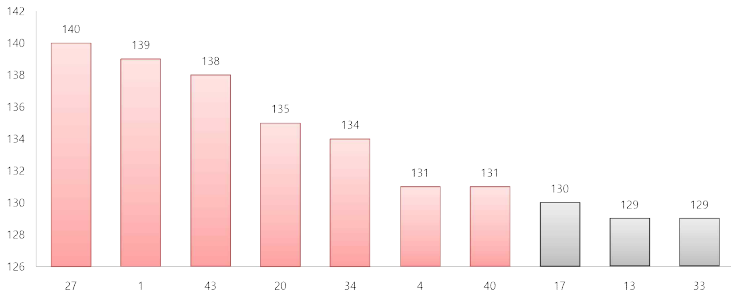
일곱번째 로또번호 분포



보너스 로또번호는
43번이 가장 많았으며, 비교적 고른 분포를 보인다

Q4. 전체 숫자에서 가장 많은 빈도를 나타낸 숫자를 나타내시오

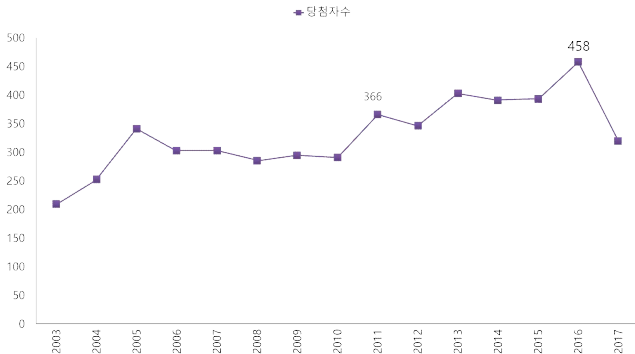
번호별 빈도수 TOP 10



전체 번호중 가장 많은 빈도수가 나타낸 번호는
27 > 1 > 43 > 20 > 34 > 4 > 40번 순이다

Q5. 연도별 당첨자수를 나타내시오

연도별 당첨자수



```
select year, sum(winner_nu) as sum_wn  
from lotto  
group by year;
```

2004년 이후 당첨자수가 300명을 넘어섰으며
2011년 이후 350명을 넘어 점차 증가하는 추세를 보임
(단, 2018년은 8월까지의 데이터)