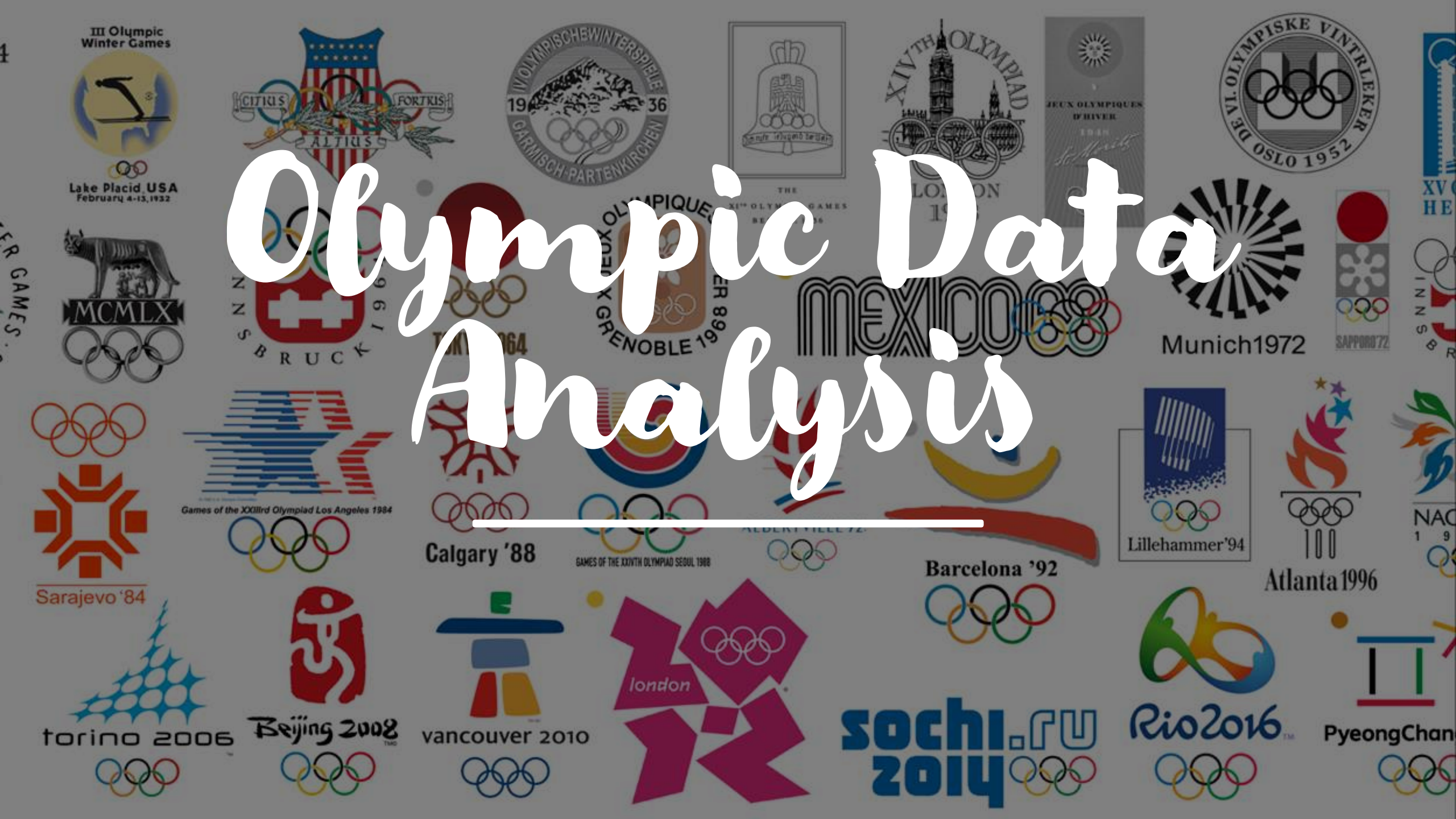


Olympic Data Analysis



Questions

Total numbers of males and females

Countries with highest number participants

The city with highest number of games played

Most number of medals by a athlete

Number of participants in each sports



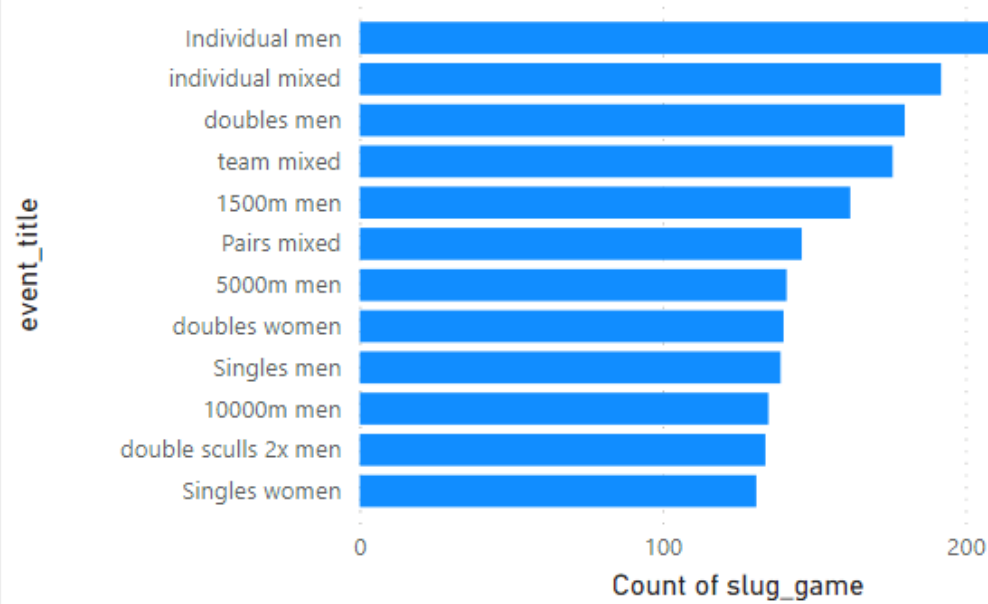
Demo in Hive



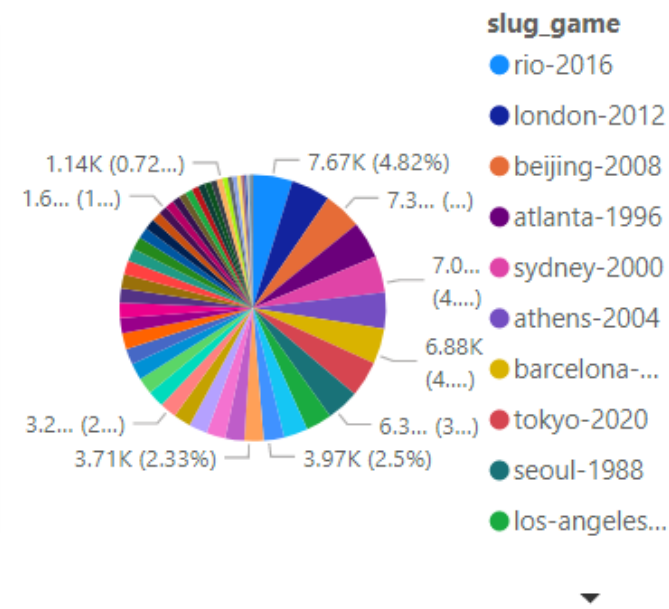
Data Visualization

Olympic Data Analysis

Number of participants in each sports



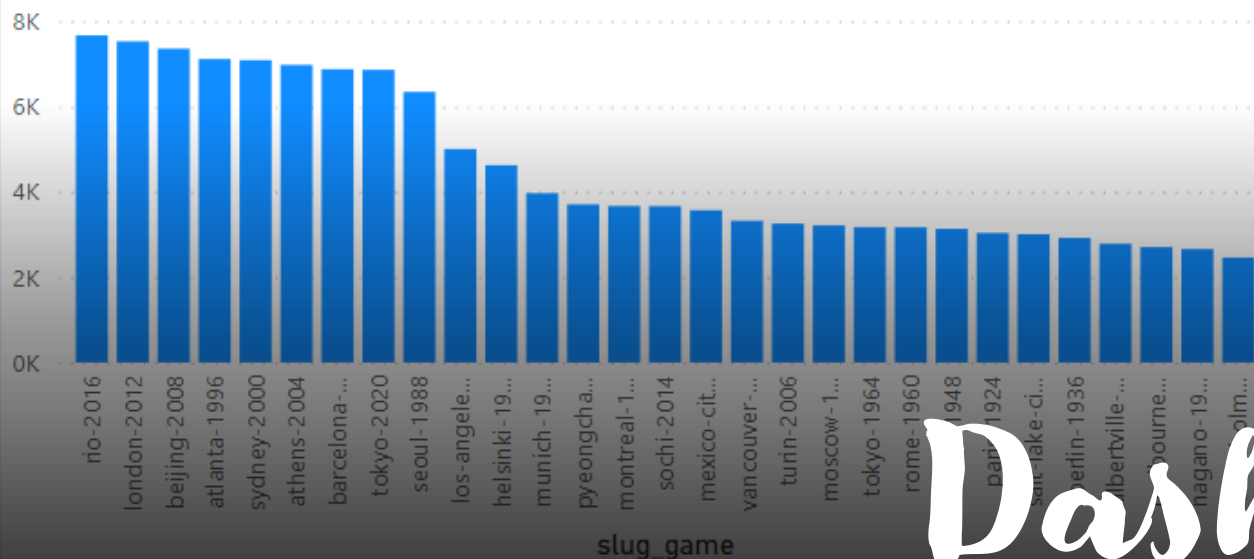
Males Vs Female Participants



Most number of medals by a athlete

athlete_full_name	Count of medal_type
Michael PHELPS	16
Larisa LATYNINA	14
#NAME?	13
Marit BJOERGEN	12
Nikolay ANDRIANOV	12
Alexei NEMOV	10
Boris SHAKHLIN	10
Takashi ONO	10
BjÅrtn DAEHLIE	9
Ireen WÅceST	9
Ole Einar BJÅRNDALLEN	9
Paavo NURMI	9
Sawao KATO	9

The city with highest number of games played



Countries with highest number participants



Dashboard

Hive Script

```
create database olympic;
```

```
use olympic;
```

```
create table medals (discipline_title String, slug_game String, event_title String, event_gender String, medal_type String, participant_type String, participant_title String, athlete_url String, country_name String, country_code String, country_3_letter_code String) row format delimited fields terminated by ',' stored as textfile tblproperties("skip.header.line.count"="1");
```

```
LOAD DATA LOCAL INPATH '/home/ubuntu/hadoop/MedalsCleaned.csv' OVERWRITE INTO TABLE medals;
```

```
create table athlete (athlete_url String, athlete_full_name String, first_game String, athlete_year_birth String, athlete_medals String, games_participations int, bio String) row format delimited fields terminated by ',' stored as textfile tblproperties("skip.header.line.count"="1");
```

```
LOAD DATA LOCAL INPATH '/home/ubuntu/hadoop/AthletesCleaned.csv' OVERWRITE INTO TABLE athlete;
```

```
create table results (discipline_title String, event_title String, slug_game String, participant_type String, medal_type String, athletes String, rank_equal String, rank_position String, country_name String, country_code String, country_3_letter_code String, athlete_url String, athlete_full_name String, value_unit String, value_type String) row format delimited fields terminated by ',' stored as textfile tblproperties("skip.header.line.count"="1");
```

```
LOAD DATA LOCAL INPATH '/home/ubuntu/hadoop/ResultsCleaned.csv' OVERWRITE INTO TABLE results;
```

Q1

```
SELECT event_gender,COUNT(*) AS Total  
FROM medals  
GROUP BY event_gender;
```

Q2

```
SELECT country_name, COUNT(*) AS medal_type  
FROM results  
GROUP BY country_name  
ORDER BY medal_type desc  
LIMIT 10;
```

Q3

```
SELECT slug_game, COUNT(*) AS Total  
FROM medals  
GROUP BY slug_game  
ORDER BY Total DESC  
LIMIT 10;
```

Q4

```
SELECT country_name, COUNT(medal_type) AS Total  
FROM medals  
GROUP BY country_name  
ORDER BY total DESC  
LIMIT 10;
```

Q5

```
SELECT event_title, COUNT(*) AS total  
FROM results  
GROUP BY event_title  
ORDER BY total DESC  
LIMIT 20;
```


The background of the image is a dense collection of small, white, cube-shaped gift boxes. Each box is decorated with a gold heart and the word "Love" written in a gold, cursive font. The boxes are scattered across a white, lace-like surface. A semi-transparent dark grey overlay covers the entire image, and the text "Thank You" is centered in a large, white, cursive font. A thin white horizontal line is positioned below the text.

Thank You
