## Use this file to document the commands and write the results you get.

#### 1. Download the data from https://drive.google.com/open?id=1TNZiSaNmcDbZ7fglKY7lPhni3ucD5Qn3

#### 2. How many rows does the dataset have?

wc cast.csv

3634468 17604656 213853188 cast.csv

#### 3. Change the delimiter to ";"

sed 's/,/;/g' cast.csv > cast14.csv

#### 4. Output the header into the separate file cast\_header.csv and remove it from the data

|  |
| --- |
| head -n 1 cast.csv > cast\_header.csv  sed –I 1d cast.csv |
|  |  |

#### 5. How many movies did Pierce Brosnan play in?

grep "Pierce Brosnan" cast.csv | wc -l

58

#### 6. Were there more superman movies or batman movies?

grep "Superman" cast.csv | wc -l

1347

grep "Batman" cast.csv | wc -l

1418

#### 7. How many roles are in this database?

cut -d "," -f 5 cast.csv | sort | uniq -c| wc -l

1509353

#### 8. What is the most often played role?

cut -d "," -f 5 cast.csv | sort | uniq -c| sort -nr | head -n 5

43297 actor

21049 actress

19025 Himself

11852 Dancer

10916 Extra

#### 9. What is the most frequent movie name?

cut -d "," -f 1 cast.csv | sort | uniq -c | sort -nr | head -n 1

1298 Around the World in Eighty Days