## 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 -l cast.csv

3634468

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

Sed “s/,/;/g” cast.csv > cast2.csv

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

Head -1 cast2.csv > cast\_header.csv

Sed -i ‘1d’ cast2.csv

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

56

Cut -d “;” -f 1-3 cast2.csv | grep “Pierce Brosnan” | sort | uniq | wc -l

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

Batman movies

Batman

Cut -d “”;” -f 1 cast2.csv | sort | uniq | grep “Superman” | wc -l

Cut -d “”;” -f 1 cast2.csv | sort | uniq | grep “Batman” | wc -l

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

3634428

Sort cast2.csv | uniq | wc -l

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

Himself

Cut -d ‘“‘ -f 2 cast2.csv | cut -d “;” -f 5 | sort | uniq -c | sort | tail

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

Hamlet

Sed “s/, /:/g” cast.csv | sed “s/,0/: /g” > clean\_titles.csv

Cut -d “,” -f 1-2 clean\_titles.csv | sort | uniq | cut -d “,” -f 1 | sort | uniq -c | sort | tail