

Lab 1

Question 2.7:

Output:

```
1935 13.0
1854 20.0
1862 22.0
1906 16.0
1893 18.0
1865 25.0
1882 20.0
[...]
1852 30.0
1863 12.0
1896 16.0
1918 32.0
1860 22.0
1870 15.0
```

Question 2.8:

Station: WILMINGTON DUPONT AP	FIPS: US	US Altitude: +0024.4
Station: CUSTOM HOUSE (CITY OFC)	FIPS: US	US Altitude: +0003.1
Station: ELIZABETH CITY MUNICIPAL AP	FIPS: US	US Altitude: +0004.0
Station: FORT CAMPBELL AAF	FIPS: US	US Altitude: +0174.7
Station: LOCKBOURNE	FIPS: US	US Altitude: +0226.8
Station: ALBANY NAVAL AIR STN	FIPS: US	US Altitude: +0066.1
Station: GREENVILLE DONALDSON AFB	FIPS: US	US Altitude: +0298.1
Station: COLUMBUS AAF	FIPS: US	US Altitude: +0066.8
Station: FORT BENNING LAWSON	FIPS: US	US Altitude: +0070.7
Station: DOTHAN REGIONAL AP	FIPS: US	US Altitude: +0107.6
Station: SAVANNAH CHATHAM FIELD AAF	FIPS: US	US Altitude: +0015.8
Station: TULLAHOMA AEDC SITE	FIPS: US	US Altitude: +0328.9
[...]		
Station: CHICAGO MEIGS FIELD	FIPS: US	US Altitude: +0180.1
Station: FRANKLIN	FIPS: US	US Altitude: +0469.1
Station: MILWAUKEE TIMMERMAN	FIPS: US	US Altitude: +0227.1
Station: DUBUQUE MUNICIPAL AP	FIPS: US	US Altitude: +0329.2
Station: WATERLOO MUNICIPAL AP	FIPS: US	US Altitude: +0267.6
Station: GRAND FORKS AF	FIPS: US	US Altitude: +0277.7
Station: HIBBING CHISHOLM-HIBBING AP	FIPS: US	US Altitude: +0413.6
Station: LINCOLN 8 ENE	FIPS: US	US Altitude: +0362.4
Station: LINCOLN 11 SW	FIPS: US	US Altitude: +0418.2
Station: TOK 70 SE	FIPS: US	US Altitude: +0609.6
Station: RUBY 44 ESE	FIPS: US	US Altitude: +0078.9

Station: SELAWIK 28 E
Station: DENALI 27 N

FIPS: US
FIPS: US
There are 30006 lines

US Altitude: +0006.7
US Altitude: +0678.2

Question 4.1:

This is how I proceeded:

- Job1: calculated word count for each words
- Job2: calculated N: total words in each docs
- Job3: calculated df: word frequency in each document
- Job4: Calculated TFIDF
- I created a separate class to calculated the words with the highest score, here is the output:

buck@callwild	108.0
myself@defoe-robinson	87.0
friday@defoe-robinson	55.0
dogs@callwild	35.0
thoughts@defoe-robinson	33.0
thornton@callwild	30.0
board@defoe-robinson	21.0
viz@defoe-robinson	20.0
towards@defoe-robinson	20.0
spitz@callwild	19.0
francois@callwild	18.0
sled@callwild	18.0
corn@defoe-robinson	18.0
afterwards@defoe-robinson	17.0
voyage@defoe-robinson	17.0
powder@defoe-robinson	16.0
providence@defoe-robinson	16.0
labor@defoe-robinson	15.0
believe@defoe-robinson	14.0
deliverance@defoe-robinson	13.0