Data analysis / manipulation (Awk)

Id	Employee Name	Job Title	Base Pay	Overtime Pay	Other Pay	Total Pay	TotalPayBenefits
1	NATHANIEL	GM	167411	0	400184	567595	567595
2	GARY	CAPTAIN	155966	245131	137811	538909	538909
3	ALBERT	CAPTAIN	212739	106088	16452	335279	335279
4	CHRISTOPHER	MECHANIC	77916	56120	198306	332343	32343
5	PATRICK	DEPUTY CHIEF	134401	9737	182234	326373	326373
6	DAVID	ASST DEPUTY	118602	8601	189082	316285	316285
7	ALSON	BATTALION CHIEF	92492	89062	134426	315981	315981
8	DAVID	DEPUTY DIRECTOR	256576	0	51322	307899	307899
10	JOANNE	CHIEF	285262	0	17115	302377	302377
12	PATRICIA	CAPTAIN	99722	87082	110804	297608	297608
13	EDWARD	EXECUTIVE	294580	0	0	294580	294580

Question i) Print EmployeeName and TotalPay who has BasePay greater than 10000

Answer

```
Id EmployeeName JobTitle        BasePay OvertimePay OtherPay TotalPay TotalPayBe
nefits
1  NATHANIEL    GM              167411  0           400184   567595   567595
2  GARY         CAPTAIN         155966  245131      137811   538909   538909
3  ALBERT       CAPTAIN         212739  106088      16452    335279   335279
4  CHRISTOPHER  MECHANIC        77916   56120       198306   332343   332343
5  PATRICK      DEPUTYCHIEF     134401  9737        182234   326373   326373
6  DAVID        ASSTDEPUTY      118602  8601        189082   316285   316285
7  ALSON        BATTALIONCHIEF  92492   89062       134426   315981   315981
8  DAVID        DEPUTYDIRECTOR  256576  0           51322    307899   307899
10 JOANNE       CHIEF           285262  0           17115    302377   302377
12 PATRICIA     CAPTAIN         99722   87082       110804   297608   297608
13 EDWARD       EXECUTIVE       294580  0           0        294580   294580

ROHIT GUPTA@ROHIT MINGW64 /e/code/linux-content (master)
$ awk '$4 > 10000 {print $2"--" $7}' data.csv
EmployeeName--TotalPay
NATHANIEL--567595
GARY--538909
ALBERT--335279
CHRISTOPHER--332343
PATRICK--326373
DAVID--316285
ALSON--315981
DAVID--307899
JOANNE--302377
PATRICIA--297608
EDWARD--294580
```

Question a) Read data file 'data.csv' from command line and extract rows which have BasePay > 10000

Answer

```
Id EmployeeName JobTitle        BasePay OvertimePay OtherPay TotalPay TotalPayBe
nefits
1  NATHANIEL    GM              167411  0           400184   567595   567595
2  GARY         CAPTAIN         155966  245131      137811   538909   538909
3  ALBERT       CAPTAIN         212739  106088      16452    335279   335279
4  CHRISTOPHER  MECHANIC        77916   56120       198306   332343   332343
5  PATRICK      DEPUTYCHIEF     134401  9737        182234   326373   326373
6  DAVID        ASSTDEPUTY      118602  8601        189082   316285   316285
7  ALSON        BATTALIONCHIEF  92492   89062       134426   315981   315981
8  DAVID        DEPUTYDIRECTOR  256576  0           51322    307899   307899
10 JOANNE       CHIEF           285262  0           17115    302377   302377
12 PATRICIA     CAPTAIN         99722   87082       110804   297608   297608
13 EDWARD       EXECUTIVE       294580  0           0        294580   294580

ROHIT GUPTA@ROHIT MINGW64 /e/code/linux-content (master)
$ awk '$4 > 10000 {print $0}' data.csv
Id EmployeeName JobTitle        BasePay OvertimePay OtherPay TotalPay TotalPayBe
nefits
1  NATHANIEL    GM              167411  0           400184   567595   567595
2  GARY         CAPTAIN         155966  245131      137811   538909   538909
3  ALBERT       CAPTAIN         212739  106088      16452    335279   335279
4  CHRISTOPHER  MECHANIC        77916   56120       198306   332343   332343
5  PATRICK      DEPUTYCHIEF     134401  9737        182234   326373   326373
6  DAVID        ASSTDEPUTY      118602  8601        189082   316285   316285
7  ALSON        BATTALIONCHIEF  92492   89062       134426   315981   315981
8  DAVID        DEPUTYDIRECTOR  256576  0           51322    307899   307899
10 JOANNE       CHIEF           285262  0           17115    302377   302377
12 PATRICIA     CAPTAIN         99722   87082       110804   297608   297608
13 EDWARD       EXECUTIVE       294580  0           0        294580   294580
```

Question b) Print only EmployeeName and TotalPay

Answer

```
Id EmployeeName JobTitle        BasePay OvertimePay OtherPay TotalPay TotalPayBenefit
s
1  NATHANIEL    GM              167411  0           400184   567595   567595
2  GARY         CAPTAIN         155966  245131      137811   538909   538909
3  ALBERT       CAPTAIN         212739  106088      16452    335279   335279
4  CHRISTOPHER  MECHANIC        77916   56120       198306   332343   332343
5  PATRICK      DEPUTYCHIEF     134401  9737        182234   326373   326373
6  DAVID        ASSTDEPUTY      118602  8601        189082   316285   316285
7  ALSON        BATTALIONCHIEF  92492   89062       134426   315981   315981
8  DAVID        DEPUTYDIRECTOR  256576  0           51322    307899   307899
10 JOANNE       CHIEF           285262  0           17115    302377   302377
12 PATRICIA     CAPTAIN         99722   87082       110804   297608   297608
13 EDWARD       EXECUTIVE       294580  0           0        294580   294580

ROHIT GUPTA@ROHIT MINGW64 /e/code/linux-content (master)
$ awk '{print $2"--"$7}' data.csv
EmployeeName--TotalPay
NATHANIEL--567595
GARY--538909
ALBERT--335279
CHRISTOPHER--332343
PATRICK--326373
DAVID--316285
ALSON--315981
DAVID--307899
JOANNE--302377
PATRICIA--297608
EDWARD--294580
```

Question ii) What is the aggregate TotalPay of employees whose jobtitle is 'CAPTAIN'

Answer

```
Id EmployeeName JobTitle         BasePay OvertimePay OtherPay TotalPay TotalPayBenefits
1  NATHANIEL    GM               167411  0           400184   567595   567595
2  GARY         CAPTAIN          155966  245131      137811   538909   538909
3  ALBERT       CAPTAIN          212739  106088      16452    335279   335279
4  CHRISTOPHER  MECHANIC         77916   56120       198306   332343   332343
5  PATRICK      DEPUTYCHIEF      134401  9737        182234   326373   326373
6  DAVID        ASSTDEPUTY       118602  8601        189082   316285   316285
7  ALSON        BATTALIONCHIEF   92492   89062       134426   315981   315981
8  DAVID        DEPUTYDIRECTOR   256576  0           51322    307899   307899
10 JOANNE       CHIEF            285262  0           17115    302377   302377
12 PATRICIA     CAPTAIN          99722   87082       110804   297608   297608
13 EDWARD       EXECUTIVE        294580  0           0        294580   294580

ROHIT GUPTA@ROHIT MINGW64 /e/code/linux-content (master)
$ awk '/CAPTAIN/{print "TotalPay of employees whose jobtitile is Captain =>"$7}' data.csv
TotalPay of employees whose jobtitile is Captain =>538909
TotalPay of employees whose jobtitile is Captain =>335279
TotalPay of employees whose jobtitile is Captain =>297608
```

Question

a) Read data file 'data.csv' from command line and extract rows which have 'CAPTAIN' in the column 'jobtitle'

Answer

```
Id EmployeeName JobTitle         BasePay OvertimePay OtherPay TotalPay TotalPayBenefits
1  NATHANIEL    GM               167411  0           400184   567595   567595
2  GARY         CAPTAIN          155966  245131      137811   538909   538909
3  ALBERT       CAPTAIN          212739  106088      16452    335279   335279
4  CHRISTOPHER  MECHANIC         77916   56120       198306   332343   332343
5  PATRICK      DEPUTYCHIEF      134401  9737        182234   326373   326373
6  DAVID        ASSTDEPUTY       118602  8601        189082   316285   316285
7  ALSON        BATTALIONCHIEF   92492   89062       134426   315981   315981
8  DAVID        DEPUTYDIRECTOR   256576  0           51322    307899   307899
10 JOANNE       CHIEF            285262  0           17115    302377   302377
12 PATRICIA     CAPTAIN          99722   87082       110804   297608   297608
13 EDWARD       EXECUTIVE        294580  0           0        294580   294580

ROHIT GUPTA@ROHIT MINGW64 /e/code/linux-content (master)
$ awk '$5 > 7000 && $5 <10000{print $3"--"$5}' data.csv
DEPUTYCHIEF--9737
ASSTDEPUTY--8601
```

Question b) Extract TotalPay and calculate sum. Print the result on terminal.

Answer

```
Id EmployeeName JobTitle         BasePay OvertimePay OtherPay TotalPay TotalPayBenefits
1  NATHANIEL    GM               167411  0           400184   567595   567595
2  GARY         CAPTAIN          155966  245131      137811   538909   538909
3  ALBERT       CAPTAIN          212739  106088      16452    335279   335279
4  CHRISTOPHER  MECHANIC         77916   56120       198306   332343   332343
5  PATRICK      DEPUTYCHIEF      134401  9737        182234   326373   326373
6  DAVID        ASSTDEPUTY       118602  8601        189082   316285   316285
7  ALSON        BATTALIONCHIEF   92492   89062       134426   315981   315981
8  DAVID        DEPUTYDIRECTOR   256576  0           51322    307899   307899
10 JOANNE       CHIEF            285262  0           17115    302377   302377
12 PATRICIA     CAPTAIN          99722   87082       110804   297608   297608
13 EDWARD       EXECUTIVE        294580  0           0        294580   294580

ROHIT GUPTA@ROHIT MINGW64 /e/code/linux-content (master)
$ awk '{print $7}{sum += $7}END{print "sum of totalpay is = " sum}' data.csv
TotalPay
567595
538909
335279
332343
326373
316285
315981
307899
302377
297608
294580
sum of totalpay is = 3935229
```

Question iii) Print JobTitle and Overtimepay who has Overtimepay is between 7000 and 10000

a) Read data file 'data.csv' from command line and extract jobtitle and overtimepay for column value range between 7000-10000

b) Print the result on terminal.

answer

```
Id EmployeeName JobTitle         BasePay OvertimePay OtherPay TotalPay TotalPayBenefits
1  NATHANIEL    GM               167411  0           400184   567595   567595
2  GARY         CAPTAIN          155966  245131      137811   538909   538909
3  ALBERT       CAPTAIN          212739  106088      16452    335279   335279
4  CHRISTOPHER  MECHANIC         77916   56120       198306   332343   332343
5  PATRICK      DEPUTYCHIEF      134401  9737        182234   326373   326373
6  DAVID        ASSTDEPUTY       118602  8601        189082   316285   316285
7  ALSON        BATTALIONCHIEF   92492   89062       134426   315981   315981
8  DAVID        DEPUTYDIRECTOR   256576  0           51322    307899   307899
10 JOANNE       CHIEF            285262  0           17115    302377   302377
12 PATRICIA     CAPTAIN          99722   87082       110804   297608   297608
13 EDWARD       EXECUTIVE        294580  0           0        294580   294580

ROHIT GUPTA@ROHIT MINGW64 /e/code/linux-content (master)
$ awk '$5 > 7000 && $5 <10000{print $3"--"$5}' data.csv
DEPUTYCHIEF--9737
ASSTDEPUTY--8601
```

Question

iv) Print average BasePay

a)Read data file 'data.csv' from command line and extract BasePay values and calculate its average
b)Print the result on terminal.

Answer

```
Id EmployeeName JobTitle        BasePay OvertimePay OtherPay TotalPay TotalPayBenefit
s
1  NATHANIEL    GM              167411  0           400184   567595   567595
2  GARY         CAPTAIN         155966  245131      137811   538909   538909
3  ALBERT       CAPTAIN         212739  106088      16452    335279   335279
4  CHRISTOPHER  MECHANIC        77916   56120       198306   332343   332343
5  PATRICK      DEPUTYCHIEF     134401  9737        182234   326373   326373
6  DAVID        ASSTDEPUTY      118602  8601        189082   316285   316285
7  ALSON        BATTALIONCHIEF  92492   89062       134426   315981   315981
8  DAVID        DEPUTYDIRECTOR  256576  0           51322    307899   307899
10 JOANNE       CHIEF           285262  0           17115    302377   302377
12 PATRICIA     CAPTAIN         99722   87082       110804   297608   297608
13 EDWARD       EXECUTIVE       294580  0           0        294580   294580

ROHIT GUPTA@DESKTOP-PMFMKKV MINGW64 /f/practice/linux-content (master)
$ awk '{print $4}{sum += $4}END{print "average of basepay is = " sum/(NR-1)}' data.csv
BasePay
167411
155966
212739
77916
134401
118602
92492
256576
285262
99722
294580
average of basepay is = 172333
```