

salaries_solutions in R

Sripad Kulkarni

11/20/2020

load the libraries

```
pacman::p_load(tidyverse, stringr)
```

load the dataset

```
sal <- read_csv("/Users/sripadk/Desktop/Salaries.csv")
head(sal)
```

```
## # A tibble: 6 x 13
##   Id EmployeeName JobTitle BasePay OvertimePay OtherPay Benefits TotalPay
##   <dbl> <chr>      <chr>      <dbl>      <dbl>      <dbl> <lgl>      <dbl>
## 1     1 NATHANIEL F~ GENERAL~ 167411.          0    400184. NA        567595.
## 2     2 GARY JIMENEZ CAPTAIN~ 155966.    245132.  137811. NA        538909.
## 3     3 ALBERT PARD~ CAPTAIN~ 212739.    106088.   16453. NA        335280.
## 4     4 CHRISTOPHER~ WIRE RO~  77916.    56121.  198307. NA        332344.
## 5     5 PATRICK GAR~ DEPUTY ~ 134402.     9737  182235. NA        326373.
## 6     6 DAVID SULLI~ ASSISTA~ 118602.     8601  189083. NA        316286.
## # ... with 5 more variables: TotalPayBenefits <dbl>, Year <dbl>, Notes <lgl>,
## #   Agency <chr>, Status <lgl>
```

Average Basepay

```
mean(sal$BasePay, na.rm = T)
```

```
## [1] 66325.45
```

What is the highest amount of OvertimePay in the dataset?

```
max(sal$OvertimePay, na.rm = T)
```

```
## [1] 245131.9
```

What is the job title of JOSEPH DRISCOLL ?

```
sal %>%
  filter(EmployeeName == 'JOSEPH DRISCOLL') %>%
  select(Id, JobTitle)
```

```
## # A tibble: 1 x 2
##   Id JobTitle
##   <dbl> <chr>
## 1    25 CAPTAIN, FIRE SUPPRESSION
```

How much does JOSEPH DRISCOLL make (including benefits)?

```
sal %>%
  filter(EmployeeName == 'JOSEPH DRISCOLL') %>%
  select(TotalPayBenefits)
```

```
## # A tibble: 1 x 1
##   TotalPayBenefits
##   <dbl>
## 1      270325.
```

What is the name of highest paid person (including benefits)?

```
sal %>%
  filter(TotalPayBenefits == max(TotalPayBenefits))
```

```
## # A tibble: 1 x 13
##   Id EmployeeName JobTitle BasePay OvertimePay OtherPay Benefits TotalPay
##   <dbl> <chr>      <chr>      <dbl>      <dbl>      <dbl> <lgl>      <dbl>
## 1     1 NATHANIEL F~ GENERAL~ 167411.          0 400184. NA      567595.
## # ... with 5 more variables: TotalPayBenefits <dbl>, Year <dbl>, Notes <lgl>,
## #   Agency <chr>, Status <lgl>
```

What is the name of lowest paid person (including benefits)?

```
sal %>%
  filter(TotalPayBenefits == min(TotalPayBenefits))
```

```
## # A tibble: 1 x 13
##   Id EmployeeName JobTitle BasePay OvertimePay OtherPay Benefits TotalPay
##   <dbl> <chr>      <chr>      <dbl>      <dbl>      <dbl> <lgl>      <dbl>
## 1 148654 Joe Lopez   Counsel~      0          0    -618. NA      -618.
## # ... with 5 more variables: TotalPayBenefits <dbl>, Year <dbl>, Notes <lgl>,
## #   Agency <chr>, Status <lgl>
```

What was the average (mean) BasePay of all employees per year? (2011-2014) ?

```
sal %>%
  group_by(Year) %>%
  summarise(mean_base_pay = mean(BasePay, na.rm = TRUE))
```

```
## # A tibble: 4 x 2
##   Year mean_base_pay
##   <dbl>         <dbl>
## 1  2011         63596.
## 2  2012         65436.
## 3  2013         69630.
## 4  2014         66564.
```

How many unique job titles are there?

```
sal %>%
  summarise(unique_title = n_distinct(JobTitle))
```

```
## # A tibble: 1 x 1
##   unique_title
##   <int>
## 1         2159
```

What are the top 5 most common jobs?

```
sal %>%
  group_by(JobTitle) %>%
  summarise(ct = n_distinct(Id)) %>%
  arrange(desc(ct)) %>%
  head(5) %>%
  ungroup()
```

```
## # A tibble: 5 x 2
##   JobTitle          ct
##   <chr>          <int>
## 1 Transit Operator    7036
## 2 Special Nurse      4389
## 3 Registered Nurse   3736
## 4 Public Svc Aide-Public Works 2518
## 5 Police Officer 3    2421
```

How many people have the word Chief in their job title?

```
sal %>%
  mutate(JobTitle = tolower(JobTitle)) %>%
  summarise(ct = sum(str_detect(JobTitle, "chief")))
```

```
## # A tibble: 1 x 1
##       ct
##   <int>
## 1    627
```

Is there a correlation between length of the Job Title string and Salary?

```
cor(str_count(sal$JobTitle), sal$TotalPayBenefits)
```

```
## [1] -0.03687845
```