

The Presence of Technical Jobs In Los Angeles

Muiz Rahemtullah

7/6/2019

Part A

Before we begin, we will first change the working directory.

```
setwd("~/Desktop/Data Science Summer 19/Project #1")
```

To find the monthly re-payment for Taylor, we will implement a function that contains the the monthly repayment formula. We will then initialize variables to the numbers that correspond with Taylor then call the function with those variables.

```
monthly_repayment_calculator <- function(loan_amount, down_payment_percentage, mortgage_intrest_rate, l
  part_1 = loan_amount * (1 - down_payment_percentage)
  part_2 = monthly_mortgage_rate * (1 + monthly_mortgage_rate)^loan_months
  part_3 = (1 + monthly_mortgage_rate)^loan_months - 1
  repayment = (part_1*part_2)/part_3
  return(repayment)
}
```

Now we will work with the values for Taylor.

```
p = 582000
d = 0.2
z = 0.045
y = 30
c = z/12
n = y*12
m = monthly_repayment_calculator(p, d, z, y, c, n)
print(m)
```

```
## [1] 2359.127
```

This figure represents the monthly repayment for Taylor.

Part B

The objective of this part is to find the proportion of technical jobs in each zipcode, and find which areas in the Greater Los Angeles Region that have the most technical jobs. We will then create a visualization on Tableau as well. ##Data Cleaning First we will load the data.

```
library(readxl)
P01_LA_zipcode_payroll <- data.frame(read_excel("P01_LA zipcode payroll.xlsx",
  sheet = "2017"))
```

Now we begin the process of cleaning the data. Firstly, we will make the Employment and Wages columns numeric columns and replace **** with 0.

```
P01_LA_zipcode_payroll$Employment <- as.numeric(P01_LA_zipcode_payroll$Employment)
```

```
## Warning: NAs introduced by coercion
```

```
P01_LA_zipcode_payroll$Wages <- as.numeric(P01_LA_zipcode_payroll$Wages)

## Warning: NAs introduced by coercion
print(class(P01_LA_zipcode_payroll$Employment))

## [1] "numeric"
print(class(P01_LA_zipcode_payroll$Wages))

## [1] "numeric"
P01_LA_zipcode_payroll$Employment[is.na(P01_LA_zipcode_payroll$Employment)] <- 0
P01_LA_zipcode_payroll$Wages[is.na(P01_LA_zipcode_payroll$Wages)] <- 0
```

Now we will make the zip code column a numeric column as well. First we will remove the total word in the column to the empty string and then convert the zip code column to a numeric column.

```
P01_LA_zipcode_payroll$Zip.Code <- gsub("Total", "", P01_LA_zipcode_payroll$Zip.Code)
P01_LA_zipcode_payroll$Zip.Code <- as.numeric(P01_LA_zipcode_payroll$Zip.Code)
print(class(P01_LA_zipcode_payroll$Zip.Code))
```

```
## [1] "numeric"
```

Finally, we will remove the Rest of the NA's.

```
P01_LA_zipcode_payroll$Industry[is.na(P01_LA_zipcode_payroll$Industry)] <- 100
P01_LA_zipcode_payroll$NAICS[is.na(P01_LA_zipcode_payroll$NAICS)] <- 100
```

Data Analysis

We will first create an output data frame with the intended results. The first is the total payroll employment.

```
output <- aggregate(P01_LA_zipcode_payroll$Employment, by = list(P01_LA_zipcode_payroll$Zip.Code, P01_L
colnames(output) = c("Zip Code", "Drop", "Total")
output$Drop <- NULL
output <- output[-c(471:7031),]
```

The aggregate function gives us the duplicate zipcodes so we cut off the rest of the data after the first iteration of zipcodes. It also creates a meaningless column so we drop that. We also rename the columns to be clear and formal.

Now we will create subsets to generate the employment in the two industries that we are interested in. We will create a subset for the Informational Job Payroll Employment and another subset for the Professional, Scientific, and Technical Skills. We then drop the extra columns and rename the remaining columns for the impending merge.

```
info <- subset(P01_LA_zipcode_payroll, P01_LA_zipcode_payroll$Industry == "Information")
professional <- subset(P01_LA_zipcode_payroll, P01_LA_zipcode_payroll$Industry == "Professional, Scienti
info <- info[, -c(2,3,4,6)]
professional <- professional[, -c(2,3,4,6)]
colnames(info) <- c("Zip Code", "Info")
colnames(professional) <- c("Zip Code", "Professional")
```

Now we will merge all datasets by the zipcode. Since we are unable to merge all three data frames with one merge, we will merge the output data frame with the other two data frames separately, and in two steps.

```
output <- merge(output, info, by = "Zip Code", all = TRUE)
output <- merge(output, professional, by = "Zip Code", all = TRUE)
```

We will add the percentage as well now.

```
output$Percentage <- (output$Info+output$Professional)/output$Total
```

Finally to conclude the project, we will create a CSV file from the final output data frame.

```
write.csv(output, "Results.csv")
```