

# Group Project

Joobee Jung

Winter 2024

```
best_case <- read.csv("best_case_final.csv")  
worst_case <- read.csv("worst_case_final.csv")
```

```
head(best_case)
```

##	CustomerID	Count	Country	State	City	Zip.Code
## 1	3668-QPYBK	1	United States	California	Los Angeles	90003
## 2	9237-HQITU	1	United States	California	Los Angeles	90005
## 3	9305-CDSKC	1	United States	California	Los Angeles	90006
## 4	7892-P00KP	1	United States	California	Los Angeles	90010
## 5	0280-XJGEX	1	United States	California	Los Angeles	90015
## 6	4190-MFLUW	1	United States	California	Los Angeles	90020
##	Lat.Long	Latitude	Longitude	Gender	Senior.Citizen	Partner
## 1	33.964131, -118.272783	33.96413	-118.2728	Male	No	No
## 2	34.059281, -118.30742	34.05928	-118.3074	Female	No	No
## 3	34.048013, -118.293953	34.04801	-118.2940	Female	No	No
## 4	34.062125, -118.315709	34.06213	-118.3157	Female	No	Yes
## 5	34.039224, -118.266293	34.03922	-118.2663	Male	No	No
## 6	34.066367, -118.309868	34.06637	-118.3099	Female	No	Yes
##	Dependents	Tenure	Phone.Service	Multiple.Lines	Internet.Service	
## 1	No	2	Yes	No	DSL	
## 2	Yes	2	Yes	No	Fiber optic	
## 3	Yes	8	Yes	Yes	Fiber optic	
## 4	Yes	28	Yes	Yes	Fiber optic	
## 5	Yes	49	Yes	Yes	Fiber optic	
## 6	No	10	Yes	No	DSL	
##	Online.Security	Online.Backup	Device.Protection	Tech.Support	Streaming.TV	
## 1	Yes	Yes	No	No	No	
## 2	No	No	No	No	No	
## 3	No	No	Yes	No	Yes	
## 4	No	No	Yes	Yes	Yes	
## 5	No	Yes	Yes	No	Yes	
## 6	No	No	Yes	Yes	No	
##	Streaming.Movies	Contract	Paperless.Billing	Payment.Method		
## 1	No Month-to-month	Yes	Mailed check			
## 2	No Month-to-month	Yes	Electronic check			
## 3	Yes Month-to-month	Yes	Electronic check			
## 4	Yes Month-to-month	Yes	Electronic check			
## 5	Yes Month-to-month	Yes	Bank transfer (automatic)			
## 6	No Month-to-month	No	Credit card (automatic)			
##	Monthly.Charges	Churn_val	Churn.Reason			
## 1	53.85	1	Competitor made better offer			
## 2	70.70	1	Moved			
## 3	99.65	1	Moved			
## 4	104.80	1	Moved			
## 5	103.70	1	Competitor had better devices			
## 6	55.20	1	Competitor offered higher download speeds			
##	churn_12month					
## 1	1					
## 2	1					
## 3	1					
## 4	0					
## 5	0					
## 6	1					

```
head(worst_case)
```

##	CustomerID	Count	Country	State	City	Zip.Code
## 1	3668-QPYBK	1	United States	California	Los Angeles	90003
## 2	9237-HQITU	1	United States	California	Los Angeles	90005
## 3	9305-CDSKC	1	United States	California	Los Angeles	90006
## 4	7892-P00KP	1	United States	California	Los Angeles	90010
## 5	0280-XJGEX	1	United States	California	Los Angeles	90015
## 6	4190-MFLUW	1	United States	California	Los Angeles	90020
##	Lat.Long	Latitude	Longitude	Gender	Senior.Citizen	Partner
## 1	33.964131, -118.272783	33.96413	-118.2728	Male	No	No
## 2	34.059281, -118.30742	34.05928	-118.3074	Female	No	No
## 3	34.048013, -118.293953	34.04801	-118.2940	Female	No	No
## 4	34.062125, -118.315709	34.06213	-118.3157	Female	No	Yes
## 5	34.039224, -118.266293	34.03922	-118.2663	Male	No	No
## 6	34.066367, -118.309868	34.06637	-118.3099	Female	No	Yes
##	Dependents	Tenure	Phone.Service	Multiple.Lines	Internet.Service	
## 1	No	2	Yes	No	DSL	
## 2	Yes	2	Yes	No	Fiber optic	
## 3	Yes	8	Yes	Yes	Fiber optic	
## 4	Yes	28	Yes	Yes	Fiber optic	
## 5	Yes	49	Yes	Yes	Fiber optic	
## 6	No	10	Yes	No	DSL	
##	Online.Security	Online.Backup	Device.Protection	Tech.Support	Streaming.TV	
## 1	Yes	Yes	No	No	No	
## 2	No	No	No	No	No	
## 3	No	No	Yes	No	Yes	
## 4	No	No	Yes	Yes	Yes	
## 5	No	Yes	Yes	No	Yes	
## 6	No	No	Yes	Yes	No	
##	Streaming.Movies	Contract	Paperless.Billing	Payment.Method		
## 1	No Month-to-month	Yes	Mailed check			
## 2	No Month-to-month	Yes	Electronic check			
## 3	Yes Month-to-month	Yes	Electronic check			
## 4	Yes Month-to-month	Yes	Electronic check			
## 5	Yes Month-to-month	Yes	Bank transfer (automatic)			
## 6	No Month-to-month	No	Credit card (automatic)			
##	Monthly.Charges	Churn_val	Churn.Reason			
## 1	53.85	1	Competitor made better offer			
## 2	70.70	1	Moved			
## 3	99.65	1	Moved			
## 4	104.80	1	Moved			
## 5	103.70	1	Competitor had better devices			
## 6	55.20	1	Competitor offered higher download speeds			
##	churn_12month					
## 1	1					
## 2	1					
## 3	1					
## 4	0					
## 5	0					
## 6	1					

```
best_case_charges <- best_case$Monthly.Charges
best_case_churn_flag <- best_case$churn_12month

worst_case_charges <- worst_case$Monthly.Charges
worst_case_churn_flag <- worst_case$churn_12month
```

# Question 1. Two-sample t-test with Monthly Charges and Customer Churn

## Best Case

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
## filter, lag
```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

```
t_test_result_best <- t.test(best_case_charges ~ best_case_churn_flag)
t_test_result_best
```

```
##
## Welch Two Sample t-test
##
## data: best_case_charges by best_case_churn_flag
## t = -2.3884, df = 1684, p-value = 0.01703
## alternative hypothesis: true difference in means between group 0 and group 1 is not
## equal to 0
## 95 percent confidence interval:
## -3.6995578 -0.3631968
## sample estimates:
## mean in group 0 mean in group 1
## 64.46260 66.49397
```

## Worst Case

```
t_test_result_worst <- t.test(worst_case_charges ~ worst_case_churn_flag)
t_test_result_worst
```

```
##
## Welch Two Sample t-test
##
## data:  worst_case_charges by worst_case_churn_flag
## t = 16.388, df = 4592, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  10.45779 13.29984
## sample estimates:
## mean in group 0 mean in group 1
##      68.31539      56.43657
```

## Question 2. Chi-Squared test with Payment Methods and Customer Churn

### Best Case

```
best_payment_contingency_table <- table(best_case_churn_flag, best_case$Payment.Method)
best_payment_contingency_table
```

```
##
## best_case_churn_flag Bank transfer (automatic) Credit card (automatic)
##              0              1441              1441
##              1              103              81
##
## best_case_churn_flag Electronic check Mailed check
##              0              1759              1365
##              1              606              247
```

```
## Payment
chisq_best_payment <- chisq.test(best_case_churn_flag, best_case$Payment.Method)
chisq_best_payment
```

```
##
## Pearson's Chi-squared test
##
## data:  best_case_churn_flag and best_case$Payment.Method
## X-squared = 411.14, df = 3, p-value < 2.2e-16
```

### Worst Case

```
worst_payment_contingency_table <- table(worst_case_churn_flag, worst_case$Payment.Method)
worst_payment_contingency_table
```

```
##
## worst_case_churn_flag Bank transfer (automatic) Credit card (automatic)
##           0           1333           1311
##           1           211           211
##
## worst_case_churn_flag Electronic check Mailed check
##           0           1414           878
##           1           951           734
```

```
## Payment
chisq_worst_payment <- chisq.test(worst_case_churn_flag, worst_case$Payment.Method)
chisq_worst_payment
```

```
##
## Pearson's Chi-squared test
##
## data:  worst_case_churn_flag and worst_case$Payment.Method
## X-squared = 688.62, df = 3, p-value < 2.2e-16
```

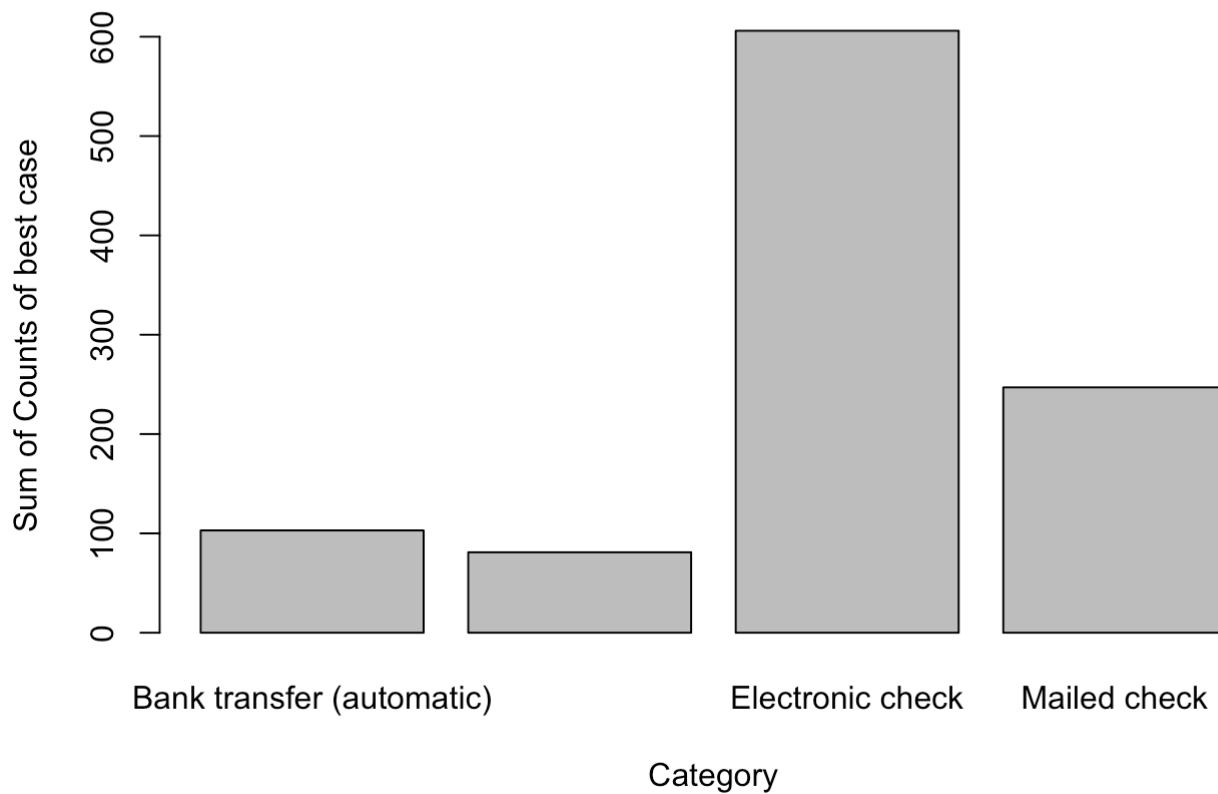
## Bar plot for payment methods

```
all_levels <- levels(factor(best_case$Payment.Method))

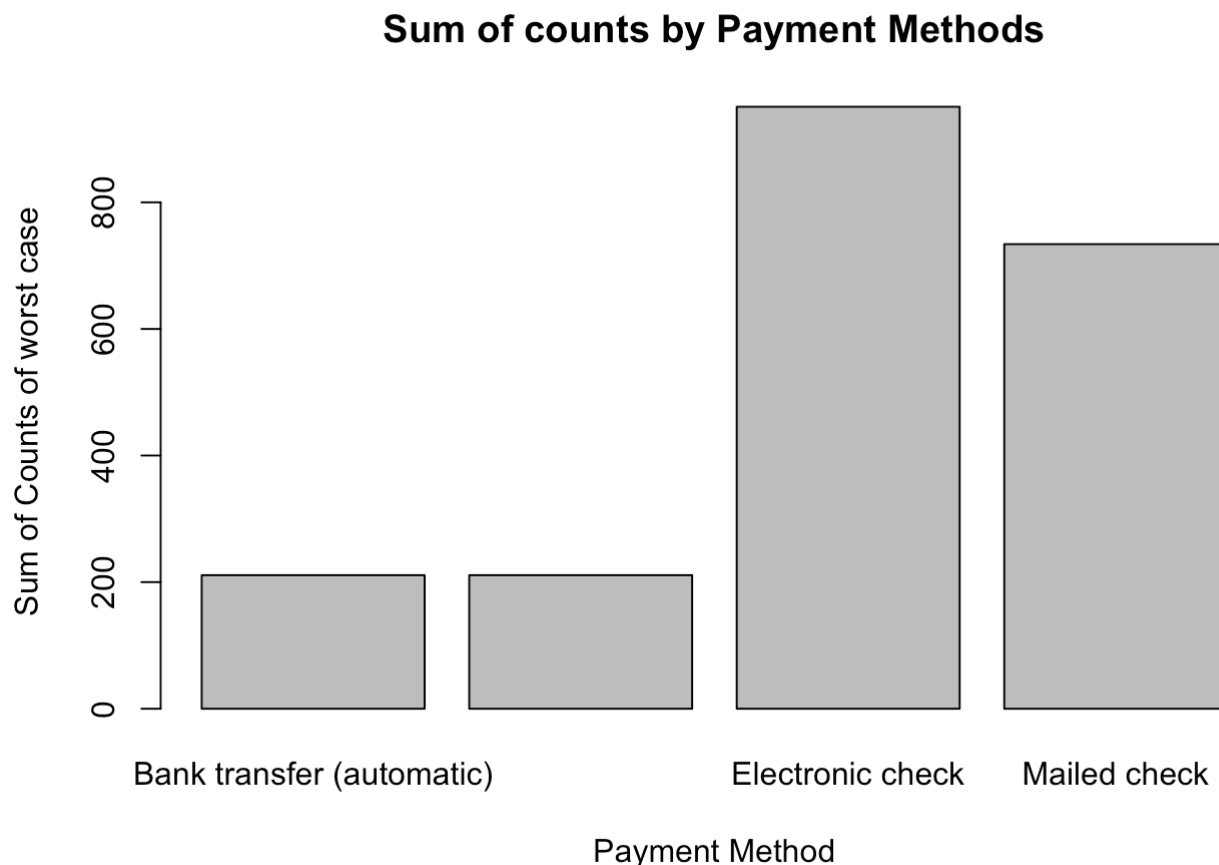
best_payment_counts <- tapply(best_case$churn_12month,
                             factor(best_case$Payment.Method, levels = all_levels),
                             sum, default = 0)

barplot(best_payment_counts,
        main = "Sum of counts by Payment Methods",
        xlab = "Category",
        ylab = "Sum of Counts of best case") # Different colors for bars
```

## Sum of counts by Payment Methods



```
worst_payment_counts <- tapply(worst_case$churn_12month,  
                               factor(worst_case$Payment.Method, levels = all_levels),  
                               sum, default = 0)  
  
barplot(worst_payment_counts,  
        main = "Sum of counts by Payment Methods",  
        xlab = "Payment Method",  
        ylab = "Sum of Counts of worst case") # Different colors for bars
```



## Question 3. Chi-Squared test with Contract types and Customer Churn

### Best case

```
best_contract_contingency_table <- table(best_case_churn_flag, best_case$Contract)
best_contract_contingency_table
```

```
##
## best_case_churn_flag Month-to-month One year Two year
##           0           2851          1460          1695
##           1           1024           13           0
```

```
## Contract
chisq_best_contract <- chisq.test(best_case_churn_flag, best_case$Contract)
chisq_best_contract
```

```
##
## Pearson's Chi-squared test
##
## data: best_case_churn_flag and best_case$Contract
## X-squared = 940.02, df = 2, p-value < 2.2e-16
```



# Worst case

```
worst_contract_contingency_table <- table(worst_case_churn_flag, worst_case$Contract)
worst_contract_contingency_table
```

```
##
## worst_case_churn_flag Month-to-month One year Two year
##           0           1934       1366       1636
##           1           1941        107         59
```

```
## Contract
chisq_worst_contract <- chisq.test(worst_case_churn_flag, worst_case$Contract)
chisq_worst_contract
```

```
##
## Pearson's Chi-squared test
##
## data:  worst_case_churn_flag and worst_case$Contract
## X-squared = 1677.7, df = 2, p-value < 2.2e-16
```

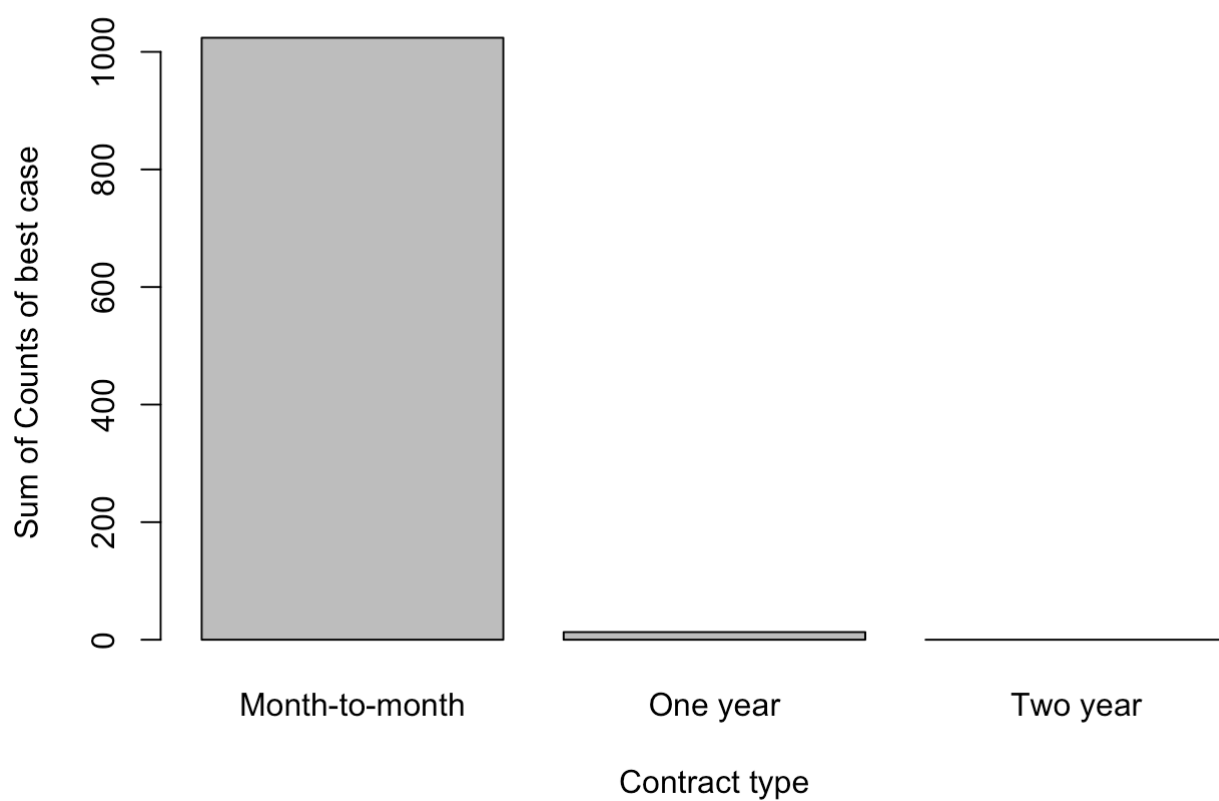
```
contract_all_levels <- levels(factor(best_case$Contract))

best_contract_counts <- tapply(best_case$churn_12month,
                              factor(best_case$Contract, levels = contract_all_levels),
                              sum, default = 0)

worst_contract_counts <- tapply(worst_case$churn_12month,
                              factor(worst_case$Contract, levels = contract_all_levels),
                              sum, default = 0)

barplot(best_contract_counts,
        main = "Sum of customer churn by contract type",
        xlab = "Contract type",
        ylab = "Sum of Counts of best case")
```

## Sum of customer churn by contract type



```
barplot(worst_contract_counts,  
        main = "Sum of customer churn by contract type",  
        xlab = "Contract type",  
        ylab = "Sum of Counts of worst case")
```

**Sum of customer churn by contract type**

