AtharvPrashantTungatkar_Individual VisualAnalytics Assignment 5

Atharv Prashant Tungatkar

2025-10-13

Read Dataset

```
siva=read.csv("C:/Users/athar/OneDrive/Desktop/MBA Business Analytics/Visual Analytics/Files/
siva.csv")

head(siva)
```

xgra_n1clb_nbr <int>	Siva_Rental_Number <int>	rent_area_loc <int>	Date_of_Survey <chr>	Day_of_W… <chr>	Time <chr>
1	51407	67041	156	5/18/2011	Wednesday	7:48::
2	23460	56084	204	2/5/2011	Saturday	22:06
3	53417	70279	181	6/14/2011	Tuesday	5:35:4
4	14382	15105	1515	1/4/2010	Monday	23:58
5	40539	49797	259	12/1/2010	Wednesday	8:24::
6	53945	71102	165	6/29/2011	Wednesday	5:34:(

6 rows | 1-8 of 30 columns

Import all the necessary libraries

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.3
```

```
##
## 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
```

Part 1 : Subsetting

SUBSET 1:

Filtering out people who gave recommendation greater than or equal to 7 and their respective bill charges and rental location.

```
subset1 <- siva %>%
  dplyr::filter(Recom_mend_Siva >= 7) %>%
  dplyr::select(Siva_Rental_Number, Recom_mend_Siva, rent_area_loc, Total_charge_USD)
head(subset1,10)
```

	Siva_Rental_Number <int>	Recom_mend_Siva <int>	rent_area_loc <int>	Total_charge_USD <dbl>
1	67041	8	156	247.29
2	56084	8	204	128.04
3	70279	8	181	75.85
4	15105	7	1515	468.51
5	49797	9	259	42.84
6	71102	9	165	107.92
7	43104	9	177	224.68
8	33940	9	2167	45.08
9	40004	9	140	130.10
10	36540	9	160	259.35

1-10 of 10 rows

SUBSET 2

Getting people who were not billed properly to get areas which are more problematic.

```
subset2 <- siva %>%
  dplyr::filter(Trans_Billing_as_Expected <= 4) %>%
  dplyr::select( Siva_Rental_Number, Trans_Billing_as_Expected, rent_area_loc, Value_for_the_
Money, Total_charge_USD)
head(subset2, 10)
```

	Siva_Rental_Number <int>	Trans_Billing_as_Expected <int>	rent_area_loc <int>	Value_for_the_Mone <int
1	12297	0	276	
2	61154	0	161	

	Siva_Rental_Number	Trans_Billing_as_Expected	rent_area_loc	Value_for_the_Mone
	<int>	<int>	<int>	<int
3	76767	1	605	
4	52242	1	161	
5	19902	3	553	
6	14029	0	160	
7	39634	1	930	
8	6706	1	161	
9	18878	4	1713	
10	5395	3	261	

SUBSET 3

Getting Average sales by weekday

```
subset3<- siva %>%
  group_by(Day_of_Week)%>%
  summarise(mean_day=mean(Total_charge_USD))
head(subset3,10)
```

Day_of_Week	mean_day
<chr>	<dbl>
Friday	245.5963
Monday	218.4416
Saturday	256.0139
Sunday	227.0171
Thursday	210.6025
Tuesday	225.8493
Wednesday	210.6890
7 rows	

SUBSET 4

Lets determine if there are significant speed differences on

different days for different purposes.

```
subset4<-siva%>%
  group_by(Purpose_of_Rental,Day_of_Week)%>%
  summarise(speed=mean(Speed_of_Service,na.rm=TRUE))
```

```
## `summarise()` has grouped output by 'Purpose_of_Rental'. You can override using
## the `.groups` argument.
```

```
head(subset4,10)
```

Purpose_of_Rental <chr>	Day_of_Week <chr>	speed <dbl>
Bus.	Friday	7.582965
Bus.	Monday	7.609482
Bus.	Saturday	7.537847
Bus.	Sunday	7.527490
Bus.	Thursday	7.535375
Bus.	Tuesday	7.541333
Bus.	Wednesday	7.545097
Ins. Rep. or Loaner	Friday	7.325581
Ins. Rep. or Loaner	Monday	7.532258
Ins. Rep. or Loaner	Saturday	6.461538

1-10 of 10 rows

SUBSET 5

Booking done using SIVA.COM

```
subset5<-siva%>%
  dplyr::filter(booking_channel_code=="SIVA.COM")%>%
  dplyr::select(Recom_mend_Siva,rent_area_loc,booking_channel_code,Purpose_of_Rental,Value_fo
r_the_Money)
head(subset5,10)
```

	Recom_mend_S… <int>	rent_area_loc <int>	booking_channel_code <chr>	Purpose_of_Rental <chr>	Value_fo
1	8	156	SIVA.COM	Bus.	
2	8	204	SIVA.COM	Leis. / Pers.	
3	8	181	SIVA.COM	Bus.	
4	7	1515	SIVA.COM	Leis. / Pers.	

Recom_mend_S...	rent_area_loc	booking_channel_code	Purpose_of_Rental	Value_fc
<int>	<int>	<chr>	<chr>	
5	9	259	SIVA.COM	Leis. / Pers.	
6	9	2167	SIVA.COM	Bus.	
7	5	953	SIVA.COM	Bus.	
8	9	140	SIVA.COM	Leis. / Pers.	
9	9	234	SIVA.COM	Leis. / Pers.	
10	9	482	SIVA.COM	Leis. / Pers.	

Part 2 : Visualizations

```
library(ggplot2)
```
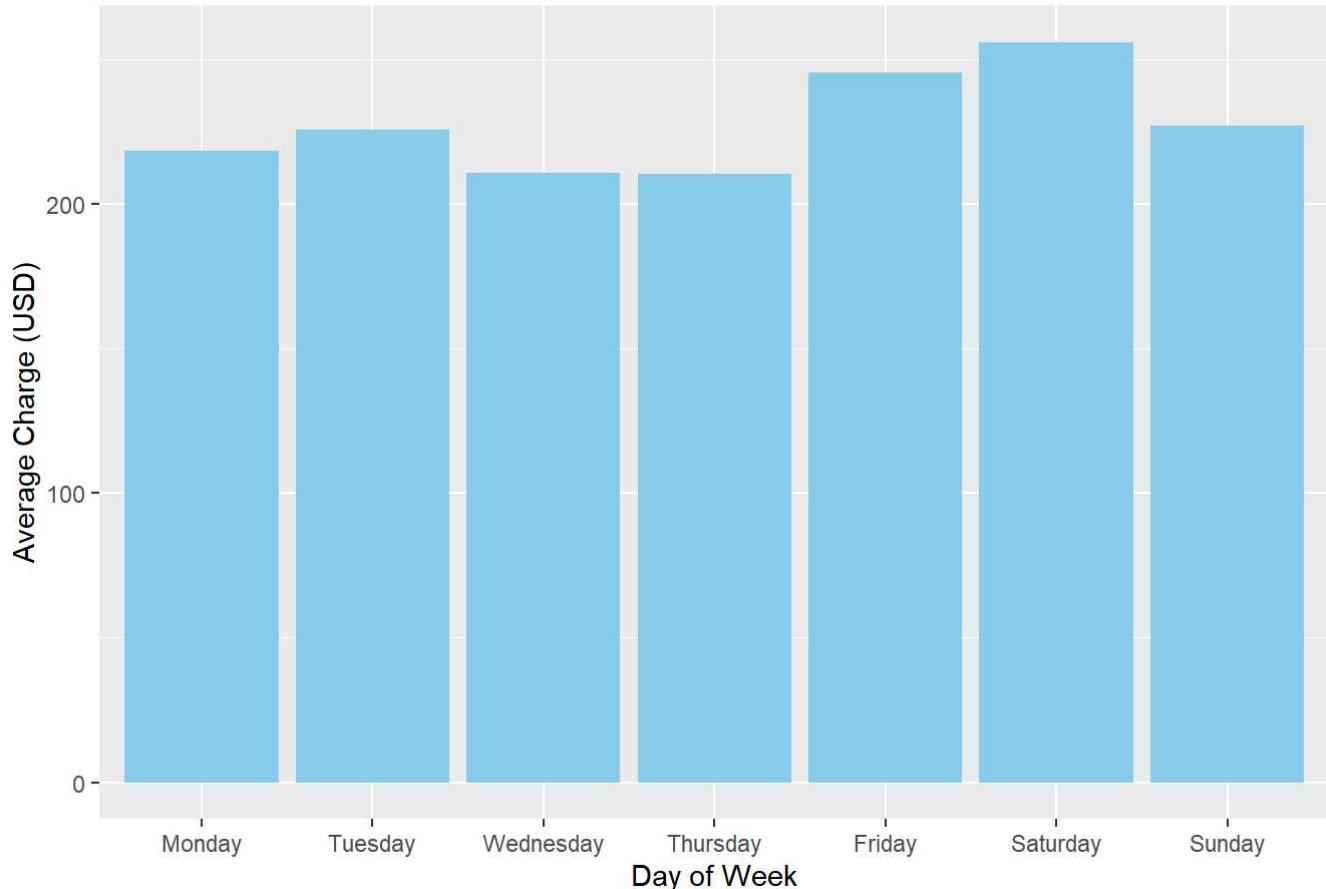
```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

Plot 1

Daywise sales

```
day_wise<- siva %>%
  group_by(Day_of_Week)%>%
  summarise(mean_day=mean(Total_charge_USD))
day_wise$Day_of_Week <- factor(day_wise$Day_of_Week,
                               levels = c("Monday", "Tuesday", "Wednesday",
                                          "Thursday", "Friday", "Saturday", "Sunday"))
ggplot(data=day_wise,aes(x=Day_of_Week,y=mean_day))+
  geom_col(fill = "skyblue") +
  labs(title = "Average Total Charge by Day of Week",
       x = "Day of Week", y = "Average Charge (USD)")
```

Average Total Charge by Day of Week

The rentals occur mostly on weekends especially Fridays and Saturdays.

Plot 2

Heatmap of Correlation between Variables

```
library(dplyr)
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.4.3
```

```
## corrplot 0.95 loaded
```

```
dat <- siva[, c("Recom_mend_Siva", "Speed_of_Service", "Value_for_the_Money", "Total_charge_U
SD","Staff_Courtesy","Veh_Equip_Condition")]
corr_matrix <- cor(dat, use = "complete.obs")
col <- colorRampPalette(c("#BB4444", "#EE9988", "#FFFFFF", "#77AADD", "#4477AA" ))

corrplot(corr_matrix)
```
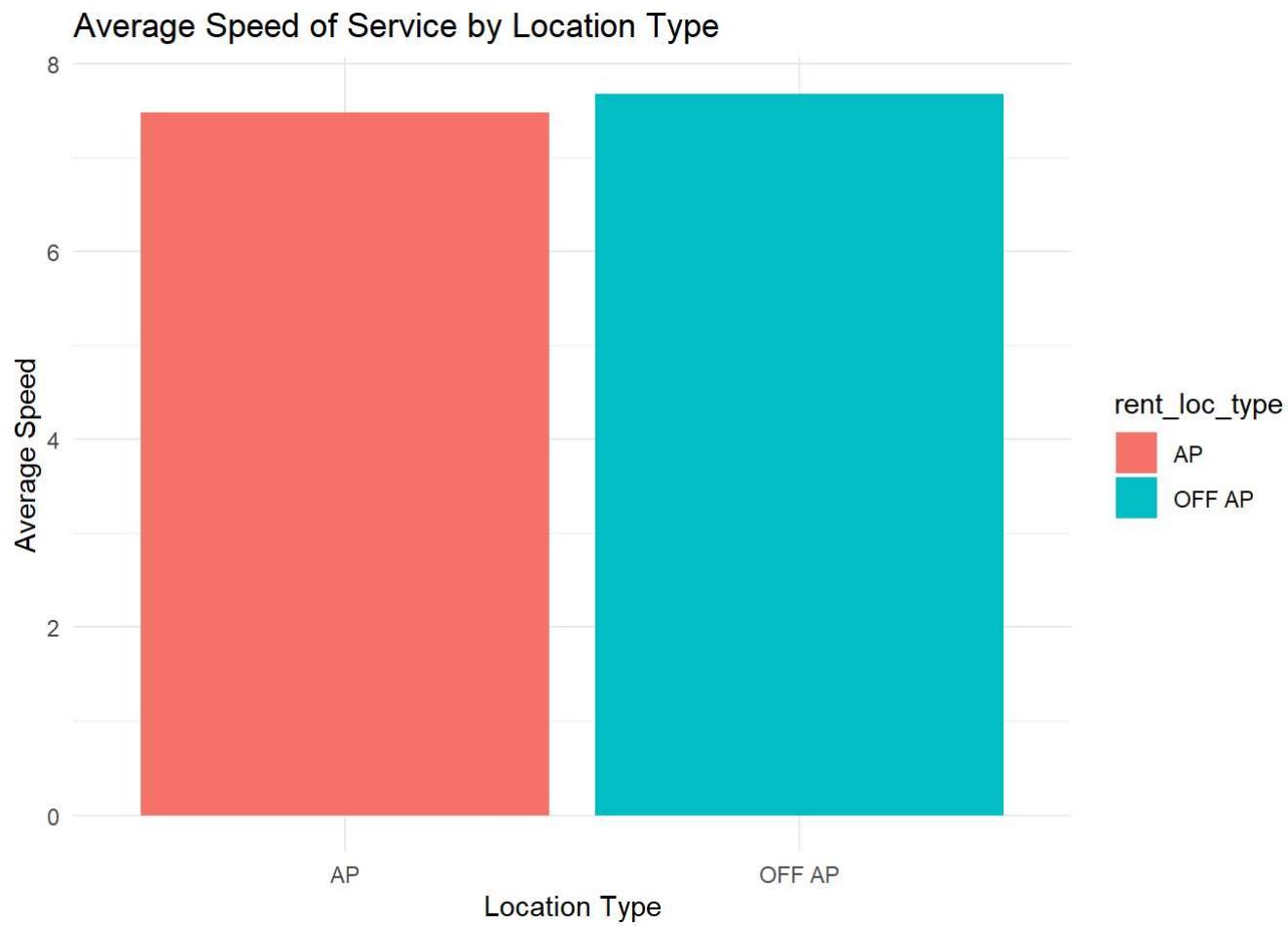
The plot reveals key information. The price is not more correlated with the services and even the recommendation doesn't depend on the price. Majority of recommendations are positively related with speed of service, staff courtesy and equipment condition.

Plot 3

Speed of serviceby location type

```
siva %>%
  filter(!is.na(rent_loc_type) & rent_loc_type != "" & rent_loc_type != "NA") %>%
  group_by(rent_loc_type) %>%
  summarise(avg_speed = mean(Speed_of_Service, na.rm = TRUE)) %>%
  ggplot(aes(x = rent_loc_type, y = avg_speed, fill = rent_loc_type)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Speed of Service by Location Type",
       x = "Location Type",
       y = "Average Speed") +
  theme_minimal()
```

Average Speed of Service by Location Type
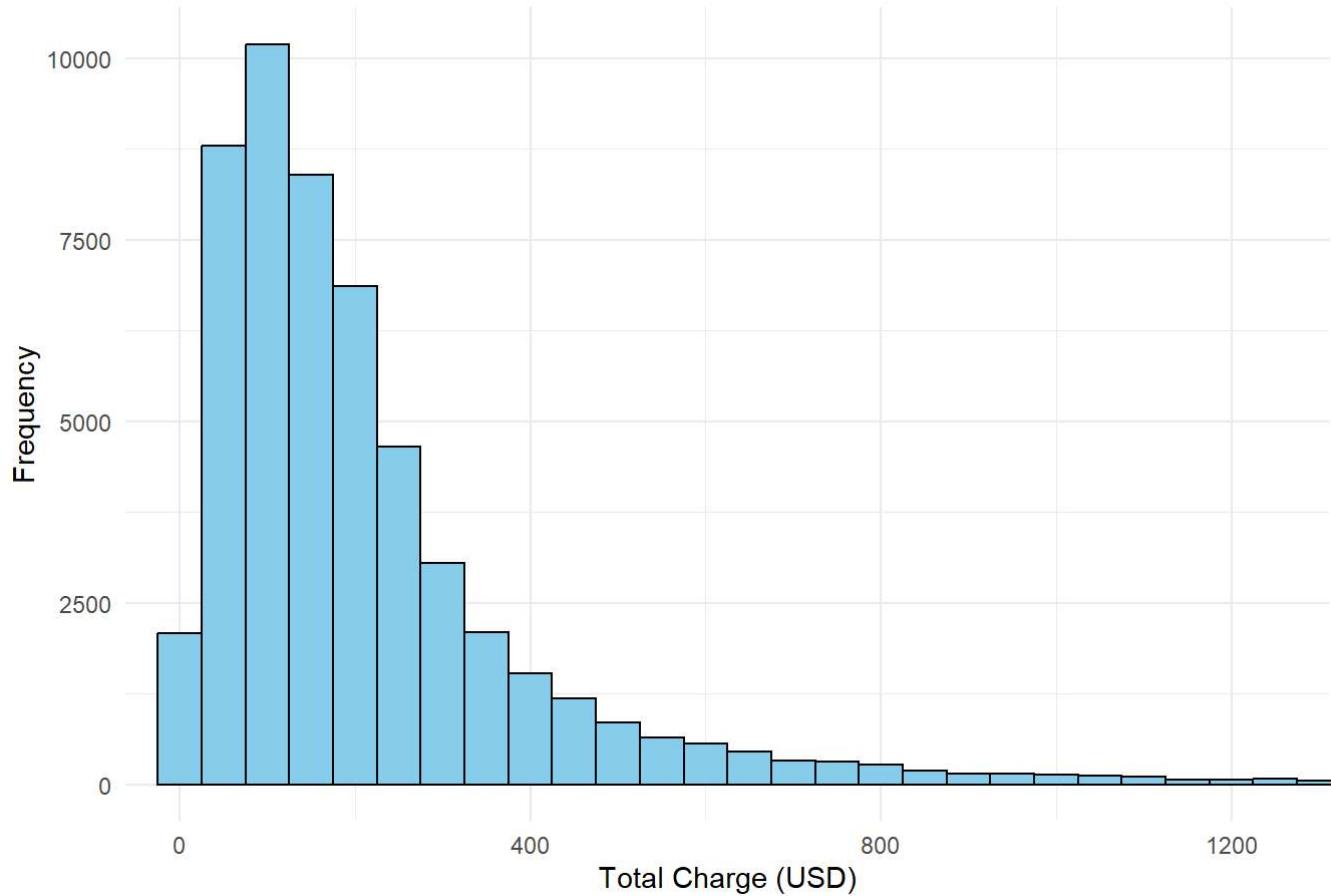
It was against my assumption that the off airport locations had a slightly higher speed.

Plot 4
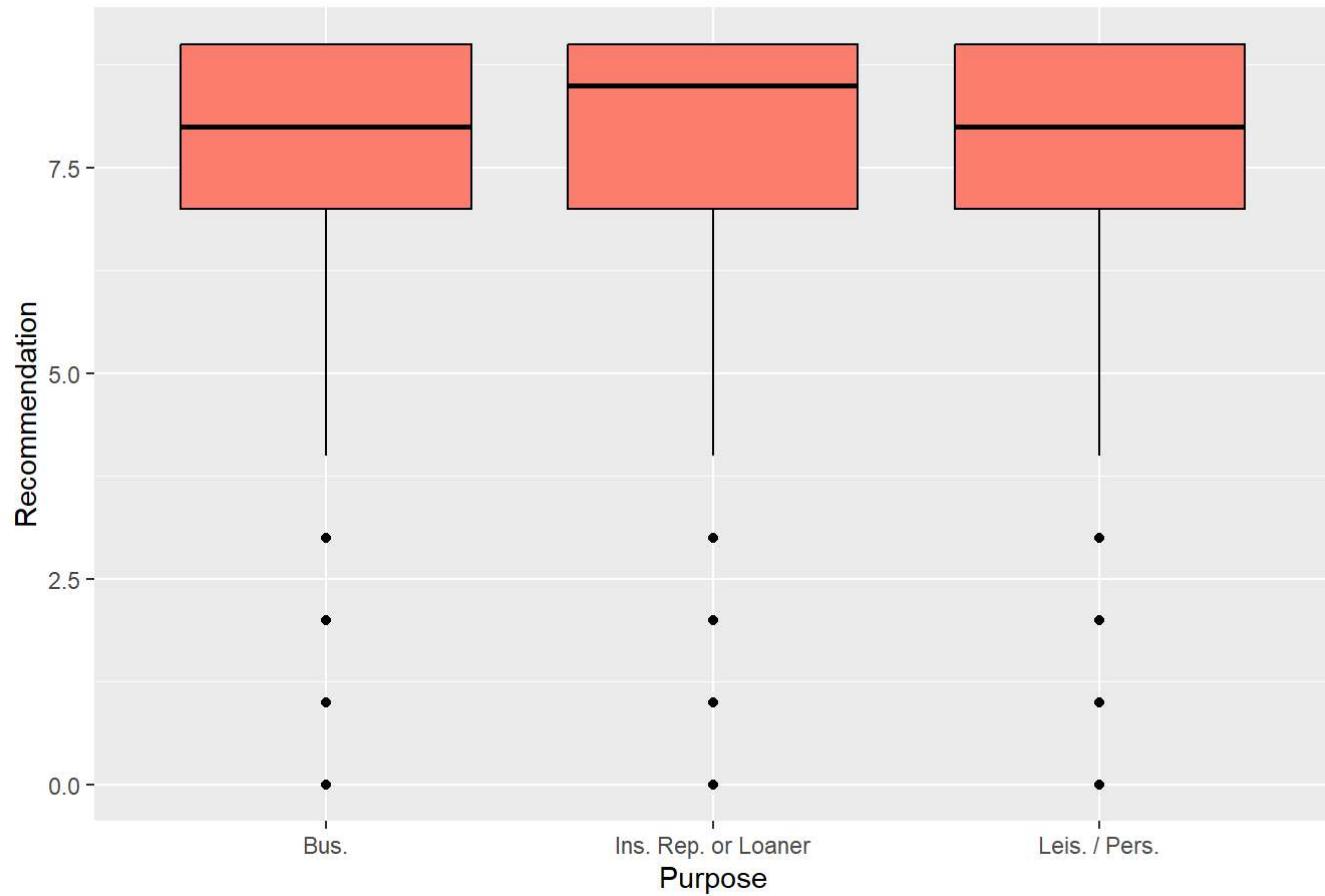
Distribiution of Total Charge

```
ggplot(siva, aes(x = Total_charge_USD)) +
  geom_histogram(binwidth = 50, fill = "skyblue", color = "black") +
  labs(title = "Distribution of Total Charges (USD)", x = "Total Charge (USD)", y = "Frequenc
y")+
  coord_cartesian(xlim = c(0, 1250)) +   # Zoom in, keeps data
  theme_minimal()
```

Distribution of Total Charges (USD)

The distribuition is unimodal and right skewed hence there are less trips that amount for more cost.

Plot 5

Box plot of Purpose of visit according to total charges

```
ggplot(siva, aes(x = Purpose_of_Rental, y =Recom_mend_Siva)) +
  geom_boxplot(fill = "salmon", color = "black") +
  labs(title = "Purpose Vs Recommendation", x = "Purpose", y = "Recommendation")
```

Purpose Vs Recommendation

Insurance replacement/loaners are more likely to recommend SIVA than other two categories.