Data Preparation:

Loading the Libraries

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.2

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.3 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ lubridate 1.9.2 ✔ tibble 3.2.1  
## ✔ purrr 1.0.2 ✔ tidyr 1.3.0  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(tigerstats)

## Loading required package: abd  
## Loading required package: nlme  
##   
## Attaching package: 'nlme'  
##   
## The following object is masked from 'package:dplyr':  
##   
## collapse  
##   
## Loading required package: lattice  
## Loading required package: grid  
## Loading required package: mosaic  
## Registered S3 method overwritten by 'mosaic':  
## method from   
## fortify.SpatialPolygonsDataFrame ggplot2  
##   
## The 'mosaic' package masks several functions from core packages in order to add   
## additional features. The original behavior of these functions should not be affected by this.  
##   
## Attaching package: 'mosaic'  
##   
## The following object is masked from 'package:Matrix':  
##   
## mean  
##   
## The following objects are masked from 'package:dplyr':  
##   
## count, do, tally  
##   
## The following object is masked from 'package:purrr':  
##   
## cross  
##   
## The following object is masked from 'package:ggplot2':  
##   
## stat  
##   
## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,  
## quantile, sd, t.test, var  
##   
## The following objects are masked from 'package:base':  
##   
## max, mean, min, prod, range, sample, sum  
##   
## Welcome to tigerstats!  
## To learn more about this package, consult its website:  
## http://homerhanumat.github.io/tigerstats

library(tibble)  
library(dplyr)  
library(stringr)  
library(SmartEDA)

## Warning: package 'SmartEDA' was built under R version 4.3.2

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(kernlab)

##   
## Attaching package: 'kernlab'  
##   
## The following object is masked from 'package:mosaic':  
##   
## cross  
##   
## The following object is masked from 'package:purrr':  
##   
## cross  
##   
## The following object is masked from 'package:ggplot2':  
##   
## alpha

library(arules)

## Warning: package 'arules' was built under R version 4.3.2

##   
## Attaching package: 'arules'  
##   
## The following object is masked from 'package:kernlab':  
##   
## size  
##   
## The following objects are masked from 'package:mosaic':  
##   
## inspect, lhs, rhs  
##   
## The following object is masked from 'package:dplyr':  
##   
## recode  
##   
## The following objects are masked from 'package:base':  
##   
## abbreviate, write

library(readxl)  
library(readr)  
library(ggpubr)

## Warning: package 'ggpubr' was built under R version 4.3.2

library(naniar)

## Warning: package 'naniar' was built under R version 4.3.2

library(MASS)

##   
## Attaching package: 'MASS'  
##   
## The following object is masked from 'package:dplyr':  
##   
## select

Loading the Dataset

data <- read.csv("C:/Users/prati/Downloads/AirlineSatisfaction/test.csv")  
head(data)

## X id Gender Customer.Type Age Type.of.Travel Class Flight.Distance  
## 1 0 19556 Female Loyal Customer 52 Business travel Eco 160  
## 2 1 90035 Female Loyal Customer 36 Business travel Business 2863  
## 3 2 12360 Male disloyal Customer 20 Business travel Eco 192  
## 4 3 77959 Male Loyal Customer 44 Business travel Business 3377  
## 5 4 36875 Female Loyal Customer 49 Business travel Eco 1182  
## 6 5 39177 Male Loyal Customer 16 Business travel Eco 311  
## Inflight.wifi.service Departure.Arrival.time.convenient  
## 1 5 4  
## 2 1 1  
## 3 2 0  
## 4 0 0  
## 5 2 3  
## 6 3 3  
## Ease.of.Online.booking Gate.location Food.and.drink Online.boarding  
## 1 3 4 3 4  
## 2 3 1 5 4  
## 3 2 4 2 2  
## 4 0 2 3 4  
## 5 4 3 4 1  
## 6 3 3 5 5  
## Seat.comfort Inflight.entertainment On.board.service Leg.room.service  
## 1 3 5 5 5  
## 2 5 4 4 4  
## 3 2 2 4 1  
## 4 4 1 1 1  
## 5 2 2 2 2  
## 6 3 5 4 3  
## Baggage.handling Checkin.service Inflight.service Cleanliness  
## 1 5 2 5 5  
## 2 4 3 4 5  
## 3 3 2 2 2  
## 4 1 3 1 4  
## 5 2 4 2 4  
## 6 1 1 2 5  
## Departure.Delay.in.Minutes Arrival.Delay.in.Minutes satisfaction  
## 1 50 44 satisfied  
## 2 0 0 satisfied  
## 3 0 0 neutral or dissatisfied  
## 4 0 6 satisfied  
## 5 0 20 satisfied  
## 6 0 0 satisfied

Observations and Attributes

num\_observations <- nrow(data)  
num\_attributes <- ncol(data)  
result <- paste("Number of Observations:", num\_observations, "\nNumber of Attributes:", num\_attributes)  
cat(result)

## Number of Observations: 25976   
## Number of Attributes: 25

Datatypes

data\_types <- sapply(data, class)  
data\_types

## X id   
## "integer" "integer"   
## Gender Customer.Type   
## "character" "character"   
## Age Type.of.Travel   
## "integer" "character"   
## Class Flight.Distance   
## "character" "integer"   
## Inflight.wifi.service Departure.Arrival.time.convenient   
## "integer" "integer"   
## Ease.of.Online.booking Gate.location   
## "integer" "integer"   
## Food.and.drink Online.boarding   
## "integer" "integer"   
## Seat.comfort Inflight.entertainment   
## "integer" "integer"   
## On.board.service Leg.room.service   
## "integer" "integer"   
## Baggage.handling Checkin.service   
## "integer" "integer"   
## Inflight.service Cleanliness   
## "integer" "integer"   
## Departure.Delay.in.Minutes Arrival.Delay.in.Minutes   
## "integer" "numeric"   
## satisfaction   
## "character"

Missing Values

missing\_values <- colSums(is.na(data))  
missing\_values

## X id   
## 0 0   
## Gender Customer.Type   
## 0 0   
## Age Type.of.Travel   
## 0 0   
## Class Flight.Distance   
## 0 0   
## Inflight.wifi.service Departure.Arrival.time.convenient   
## 0 0   
## Ease.of.Online.booking Gate.location   
## 0 0   
## Food.and.drink Online.boarding   
## 0 0   
## Seat.comfort Inflight.entertainment   
## 0 0   
## On.board.service Leg.room.service   
## 0 0   
## Baggage.handling Checkin.service   
## 0 0   
## Inflight.service Cleanliness   
## 0 0   
## Departure.Delay.in.Minutes Arrival.Delay.in.Minutes   
## 0 83   
## satisfaction   
## 0

Data Cleaning:

library(dplyr)  
# Exclude the missing observations  
data <-data %>%  
na.omit()   
dim(data)

## [1] 25893 25

missing\_values <- colSums(is.na(data))  
missing\_values

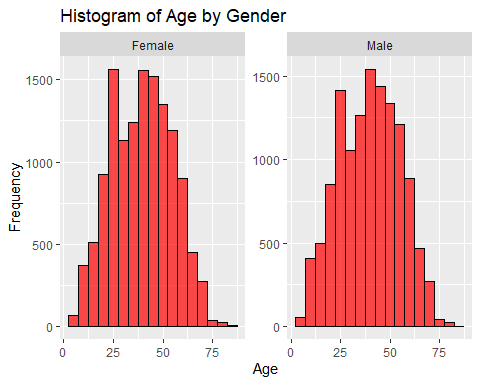
## X id   
## 0 0   
## Gender Customer.Type   
## 0 0   
## Age Type.of.Travel   
## 0 0   
## Class Flight.Distance   
## 0 0   
## Inflight.wifi.service Departure.Arrival.time.convenient   
## 0 0   
## Ease.of.Online.booking Gate.location   
## 0 0   
## Food.and.drink Online.boarding   
## 0 0   
## Seat.comfort Inflight.entertainment   
## 0 0   
## On.board.service Leg.room.service   
## 0 0   
## Baggage.handling Checkin.service   
## 0 0   
## Inflight.service Cleanliness   
## 0 0   
## Departure.Delay.in.Minutes Arrival.Delay.in.Minutes   
## 0 0   
## satisfaction   
## 0

Data Exploration (EDA):

Data Analysis

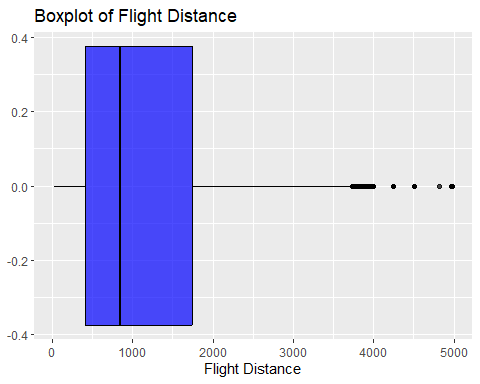
Histogram

ggplot(data, aes(x = Age, fill = Gender)) +  
 geom\_histogram(binwidth = 5, fill= "red",color = "black", alpha = 0.7) +  
 labs(title = "Histogram of Age by Gender",  
 x = "Age",  
 y = "Frequency") +  
 facet\_wrap(~Gender, scales = "free")

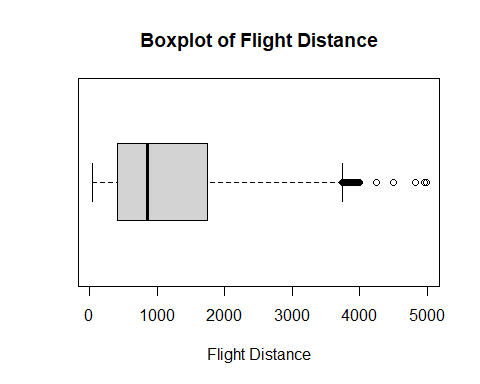


Boxplot

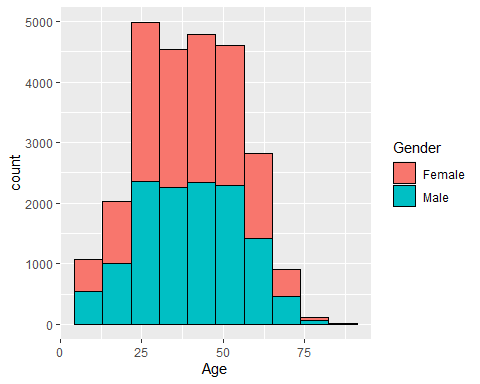
ggplot(data, aes(x = Flight.Distance)) +  
 geom\_boxplot(fill = "blue", color = "black", alpha = 0.7) +  
 labs(title = "Boxplot of Flight Distance",  
 x = "Flight Distance")



boxplot(data$Flight.Distance, main = "Boxplot of Flight Distance",  
 xlab = "Flight Distance", horizontal = TRUE)

 Overlaying Histogram

ggplot(data, aes(x = Age,fill=Gender)) +  
geom\_histogram(aes(fill = Gender),bins = 10, color="black")



dim(data)

## [1] 25893 25

summary(data)

## X id Gender Customer.Type   
## Min. : 0 Min. : 17 Length:25893 Length:25893   
## 1st Qu.: 6496 1st Qu.: 32209 Class :character Class :character   
## Median :12984 Median : 65344 Mode :character Mode :character   
## Mean :12988 Mean : 65022   
## 3rd Qu.:19482 3rd Qu.: 97623   
## Max. :25975 Max. :129877   
## Age Type.of.Travel Class Flight.Distance  
## Min. : 7.00 Length:25893 Length:25893 Min. : 31   
## 1st Qu.:27.00 Class :character Class :character 1st Qu.: 414   
## Median :40.00 Mode :character Mode :character Median : 849   
## Mean :39.62 Mean :1194   
## 3rd Qu.:51.00 3rd Qu.:1744   
## Max. :85.00 Max. :4983   
## Inflight.wifi.service Departure.Arrival.time.convenient Ease.of.Online.booking  
## Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000   
## Median :3.000 Median :3.000 Median :3.000   
## Mean :2.724 Mean :3.046 Mean :2.756   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :5.000 Max. :5.000 Max. :5.000   
## Gate.location Food.and.drink Online.boarding Seat.comfort   
## Min. :1.000 Min. :0.000 Min. :0.000 Min. :1.000   
## 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000   
## Median :3.000 Median :3.000 Median :4.000 Median :4.000   
## Mean :2.976 Mean :3.215 Mean :3.262 Mean :3.449   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:5.000   
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000   
## Inflight.entertainment On.board.service Leg.room.service Baggage.handling  
## Min. :0.000 Min. :0.000 Min. :0.00 Min. :1.000   
## 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.00 1st Qu.:3.000   
## Median :4.000 Median :4.000 Median :4.00 Median :4.000   
## Mean :3.357 Mean :3.386 Mean :3.35 Mean :3.633   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.00 3rd Qu.:5.000   
## Max. :5.000 Max. :5.000 Max. :5.00 Max. :5.000   
## Checkin.service Inflight.service Cleanliness Departure.Delay.in.Minutes  
## Min. :1.000 Min. :0.000 Min. :0.000 Min. : 0.00   
## 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:2.000 1st Qu.: 0.00   
## Median :3.000 Median :4.000 Median :3.000 Median : 0.00   
## Mean :3.314 Mean :3.649 Mean :3.286 Mean : 14.23   
## 3rd Qu.:4.000 3rd Qu.:5.000 3rd Qu.:4.000 3rd Qu.: 12.00   
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :1128.00   
## Arrival.Delay.in.Minutes satisfaction   
## Min. : 0.00 Length:25893   
## 1st Qu.: 0.00 Class :character   
## Median : 0.00 Mode :character   
## Mean : 14.74   
## 3rd Qu.: 13.00   
## Max. :1115.00

Hypothesis Testing

random\_sample <- data %>%  
 sample\_n(50)  
  
head(random\_sample, 20)

## X id Gender Customer.Type Age Type.of.Travel Class  
## 1 5199 118414 Male Loyal Customer 35 Business travel Business  
## 2 11095 79263 Male Loyal Customer 7 Personal Travel Eco  
## 3 13876 8107 Female Loyal Customer 29 Business travel Business  
## 4 20912 13449 Female Loyal Customer 26 Personal Travel Eco  
## 5 13298 12888 Male Loyal Customer 46 Business travel Business  
## 6 24463 45322 Male Loyal Customer 17 Business travel Eco  
## 7 23023 52436 Male Loyal Customer 8 Personal Travel Eco  
## 8 20326 114101 Female Loyal Customer 49 Business travel Eco Plus  
## 9 7229 18147 Female Loyal Customer 66 Personal Travel Business  
## 10 23239 55609 Female disloyal Customer 17 Business travel Eco  
## 11 6491 54616 Male disloyal Customer 42 Business travel Eco  
## 12 16818 28556 Female Loyal Customer 20 Personal Travel Eco  
## 13 8222 41405 Female Loyal Customer 29 Business travel Business  
## 14 7206 92125 Male Loyal Customer 21 Personal Travel Eco  
## 15 25009 5225 Male Loyal Customer 32 Business travel Business  
## 16 3013 15601 Female disloyal Customer 54 Business travel Eco  
## 17 10046 70459 Male Loyal Customer 60 Business travel Business  
## 18 10846 4379 Male disloyal Customer 25 Business travel Business  
## 19 13408 109796 Male disloyal Customer 27 Business travel Business  
## 20 7132 28132 Male disloyal Customer 27 Business travel Eco  
## Flight.Distance Inflight.wifi.service Departure.Arrival.time.convenient  
## 1 2422 5 5  
## 2 1598 3 2  
## 3 402 3 3  
## 4 491 2 5  
## 5 3391 2 2  
## 6 188 5 5  
## 7 651 2 3  
## 8 447 5 1  
## 9 306 4 4  
## 10 404 2 2  
## 11 854 1 1  
## 12 2701 1 4  
## 13 2222 2 2  
## 14 1589 3 5  
## 15 2434 2 1  
## 16 229 2 0  
## 17 3858 3 3  
## 18 473 3 5  
## 19 1011 4 3  
## 20 550 3 3  
## Ease.of.Online.booking Gate.location Food.and.drink Online.boarding  
## 1 5 5 2 5  
## 2 3 4 3 3  
## 3 3 3 5 5  
## 4 2 2 1 2  
## 5 2 2 4 2  
## 6 3 5 5 5  
## 7 2 5 1 2  
## 8 1 1 2 4  
## 9 3 3 2 3  
## 10 1 4 1 1  
## 11 1 2 4 1  
## 12 1 3 2 1  
## 13 2 2 2 2  
## 14 3 3 4 3  
## 15 5 1 2 2  
## 16 2 4 2 1  
## 17 4 3 2 5  
## 18 3 2 4 3  
## 19 3 2 4 3  
## 20 3 3 2 3  
## Seat.comfort Inflight.entertainment On.board.service Leg.room.service  
## 1 4 5 5 5  
## 2 3 3 1 3  
## 3 5 5 4 3  
## 4 1 1 4 5  
## 5 1 5 5 5  
## 6 5 5 3 5  
## 7 1 1 2 1  
## 8 1 5 5 5  
## 9 4 5 5 3  
## 10 1 1 2 3  
## 11 4 4 1 1  
## 12 2 2 4 5  
## 13 2 2 2 3  
## 14 4 4 3 3  
## 15 2 2 3 3  
## 16 1 1 5 4  
## 17 5 4 4 4  
## 18 4 4 5 5  
## 19 4 4 3 4  
## 20 2 2 4 2  
## Baggage.handling Checkin.service Inflight.service Cleanliness  
## 1 5 5 5 5  
## 2 3 1 4 3  
## 3 4 4 4 5  
## 4 3 2 3 1  
## 5 5 5 5 3  
## 6 2 4 5 5  
## 7 3 5 5 1  
## 8 5 3 5 4  
## 9 3 4 5 3  
## 10 3 4 4 1  
## 11 4 1 5 4  
## 12 4 5 4 2  
## 13 4 4 3 2  
## 14 4 4 4 4  
## 15 4 1 4 2  
## 16 4 1 3 1  
## 17 4 4 4 3  
## 18 1 5 2 4  
## 19 4 5 4 4  
## 20 1 2 5 2  
## Departure.Delay.in.Minutes Arrival.Delay.in.Minutes satisfaction  
## 1 29 23 satisfied  
## 2 0 0 neutral or dissatisfied  
## 3 0 0 satisfied  
## 4 0 27 neutral or dissatisfied  
## 5 0 11 satisfied  
## 6 0 0 satisfied  
## 7 0 0 neutral or dissatisfied  
## 8 4 0 satisfied  
## 9 0 0 neutral or dissatisfied  
## 10 34 9 neutral or dissatisfied  
## 11 0 0 neutral or dissatisfied  
## 12 17 0 neutral or dissatisfied  
## 13 13 21 neutral or dissatisfied  
## 14 0 0 neutral or dissatisfied  
## 15 4 5 neutral or dissatisfied  
## 16 131 140 neutral or dissatisfied  
## 17 0 0 satisfied  
## 18 4 2 neutral or dissatisfied  
## 19 35 22 satisfied  
## 20 0 0 neutral or dissatisfied

T-Test

t.test(random\_sample$Online.boarding, mu = mean(data$Online.boarding))

##   
## One Sample t-test  
##   
## data: random\_sample$Online.boarding  
## t = -0.5168, df = 49, p-value = 0.6076  
## alternative hypothesis: true mean is not equal to 3.261615  
## 95 percent confidence interval:  
## 2.764868 3.555132  
## sample estimates:  
## mean of x   
## 3.16

Linear Regression

library(ggplot2)  
  
  
lm\_model <- lm(Cleanliness ~ Checkin.service, data = data)  
  
#here we created a scatter plot with regression line  
ggplot(data, aes(x = Checkin.service, y = Cleanliness)) +  
 geom\_point() + # Scatter plot points  
 geom\_smooth(method = "lm", # Add linear regression line  
 se = FALSE, # Don't show confidence interval  
 col = "blue") + # Line color  
 labs(title = "Linear Regression: Checkin.service vs Cleanliness",  
 x = "Checkin.service",  
 y = "Cleanliness") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'

