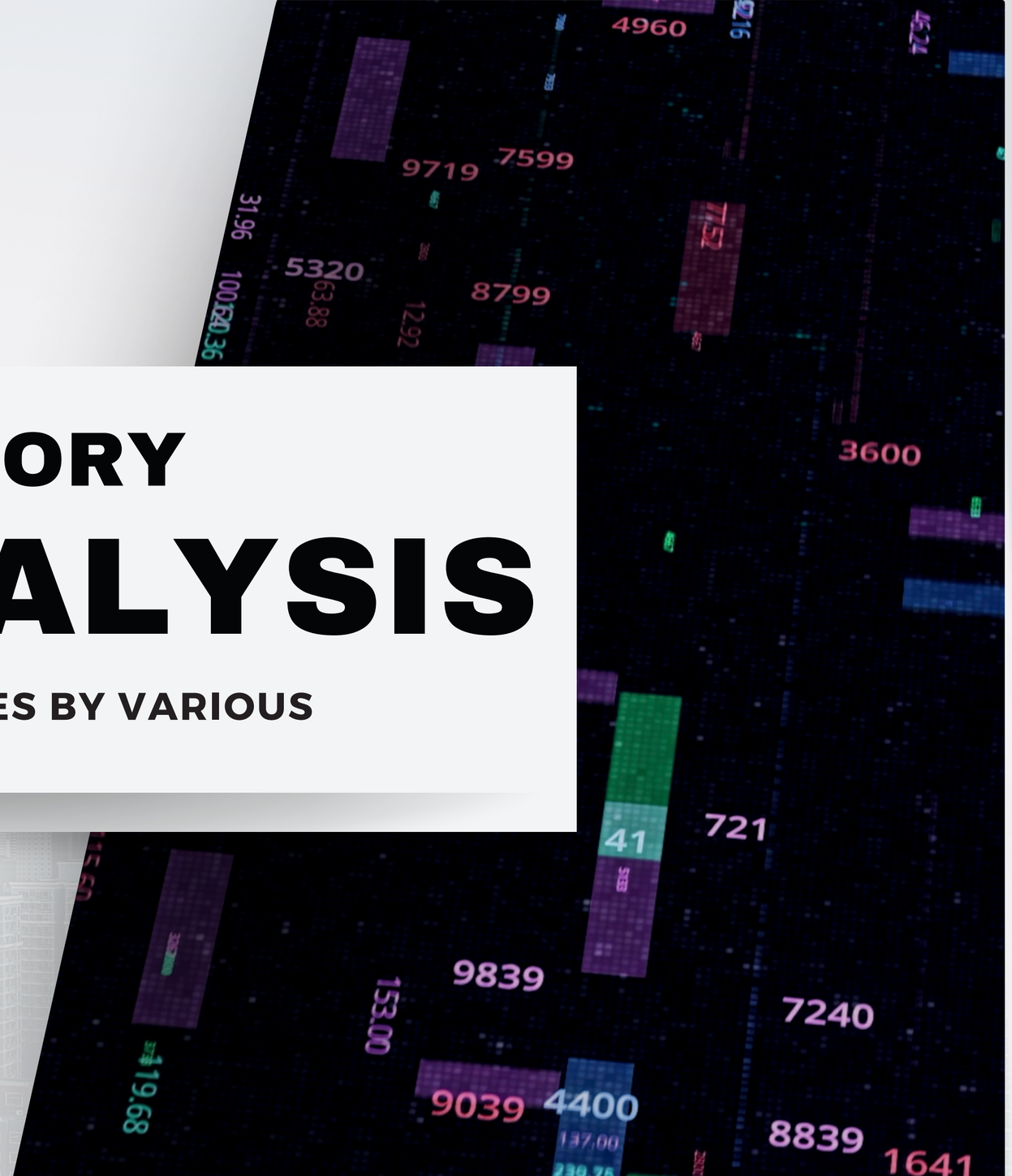
EXPLORATORY
DATA ANALYSIS

ANALYSIS OF EMPLOYEE SALARIES BY VARIOUS FACTORS



ALISHBA RIZWAN | JULY 2024

WWW.LINKEDIN.COM/IN/ALISHBA-RIZWAN--

INTRODUCTION

OBJECTIVE

To analyze and visualize salary data to uncover patterns and insights.

DATA SOURCE

The dataset includes employee details such as names, gender, dates of joining, current dates, designations, ages, salaries, units, leaves used, leaves remaining, ratings, and past experience.

- Observations: 2,639
- Variables: 13 variables

METHODS

- Data Cleaning
- Exploratory Data Analysis (EDA)
- Visualizations

DATASET

Filter

	FIRST.NAME	LAST.NAME	SEX	DOJ	CURRENT.DATE	DESIGNATION	AGE	SALARY	UNIT	LEAVES.USED	LEAVES.REMAINING	RATINGS	PAST.EXP	Tenure
1	TOMASA	ARMEN	F	2014-05-18	2016-01-07	Analyst	21	44570	Finance	24	6	2	0	1.6399726
3	OLIVE	ANCY	F	2014-07-28	2016-01-07	Analyst	21	40955	Finance	23	7	3	0	1.4455852
4	CHERRY	AQUILAR	F	2013-04-03	2016-01-07	Analyst	22	45550	IT	22	8	3	0	2.7624914
5	LEON	ABOULAHOUD	M	2014-11-20	2016-01-07	Analyst	24	43161	Operations	27	3	3	3	1.1307324
6	VICTORIA		F	2013-02-19	2016-01-07	Analyst	22	48736	Marketing	20	10	4	0	2.8802190
7	ELLIOT	AGULAR	M	2013-09-02	2016-01-07	Analyst	22	40339	Marketing	19	11	5	0	2.3463381
8	JACQUES	AKMAL	M	2013-12-05	2016-01-07	Analyst	24	40058	Marketing	29	1	2	2	2.0889802
9	KATHY	ALSOP	F	2014-06-29	2016-01-07	Senior Analyst	28	63478	Operations	20	10	3	1	1.5249829
10	LILIAN	APELA	F	2014-11-11	2016-01-07	Analyst	22	43110	Finance	15	15	3	0	1.1553730
11	BELLE	ARDS	F	2014-03-10	2016-01-07	Analyst	24	41590	Marketing	22	8	4	1	1.8288843
12	VIRGIL	ACKIES	M	2010-02-01	2016-01-07	Senior Manager	36	160613	Finance	22	11	2	9	5.9301848
13	WELDON	AIVAO	M	2013-08-01	2016-01-07	Analyst	24	44828	Finance	15	15	5	1	2.4339493
14	BOYD	AFTON	M	2013-03-22	2016-01-07	Analyst	21	45830	Web	23	7	2	0	2.7953457
15	BART	AGUILLERA	M	2013-07-27	2016-01-07	Analyst	24	43457	Management	30	0	4	1	2.4476386
16	CORINNE	ANDRZEJCZYK	F	2014-08-13	2016-01-07	Analyst	21	44812	IT	16	8	5	0	1.4017796
17	ALONZO	ADSIDE	M	2013-09-16	2016-01-07	Analyst	25	47636	Finance	27	3	3	2	2.3080082
18	ROYCE	AGOSTO	M	2014-06-07	2016-01-07	Analyst	24	48651	Web	27	3	5	1	1.5852156
19	BURTON	AGUILER	M	2013-02-03	2016-01-07	Analyst	23	40411	Management	21	9	3	0	2.9240246
20	PHILLIP	ABDULA	M	2013-07-03	2016-01-07	Analyst	24	44665	Web	22	12	5	0	2.5133470

Showing 1 to 19 of 2,638 entries, 14 total columns

DATA CLEANING

Missing Values Handling:

- Replaced missing AGE values with the median age of the dataset.
- Removed rows with excessive missing data, such as the row containing Annie’s details.

DATA CLEANING

```r
median_age <- median(salary_data$AGE, na.rm = TRUE)
salary_data$AGE[is.na(salary_data$AGE)] <- median_age

median_leavesused <- median(salary_data$LEAVES.USED , na.rm = TRUE)
salary_data$LEAVES.USED [is.na(salary_data$LEAVES.USED)] <-median_leavesused

median_leavesremaining <- median(salary_data$LEAVES.REMAINING , na.rm = TRUE)
salary_data$LEAVES.REMAINING [is.na(salary_data$LEAVES.REMAINING)] <-median_leavesrem

median_ratings <- median(salary_data$RATINGS , na.rm = TRUE)
salary_data$RATINGS [is.na(salary_data$RATINGS)] <-median_ratings

salary_data$SALARY <- as.numeric(salary_data$SALARY)

summary(salary_data)


salary_data$DOJ <- as.Date(salary_data$DOJ, format = "%m-%d-%Y")
salary_data$CURRENT.DATE <- as.Date(salary_data$CURRENT.DATE, format = "%m-%d-%Y")
salary_data$UNIT <- as.factor(salary_data$UNIT)
salary_data <- salary_data[salary_data$FIRST.NAME != "ANNIE", ]
summary(salary_data)
```

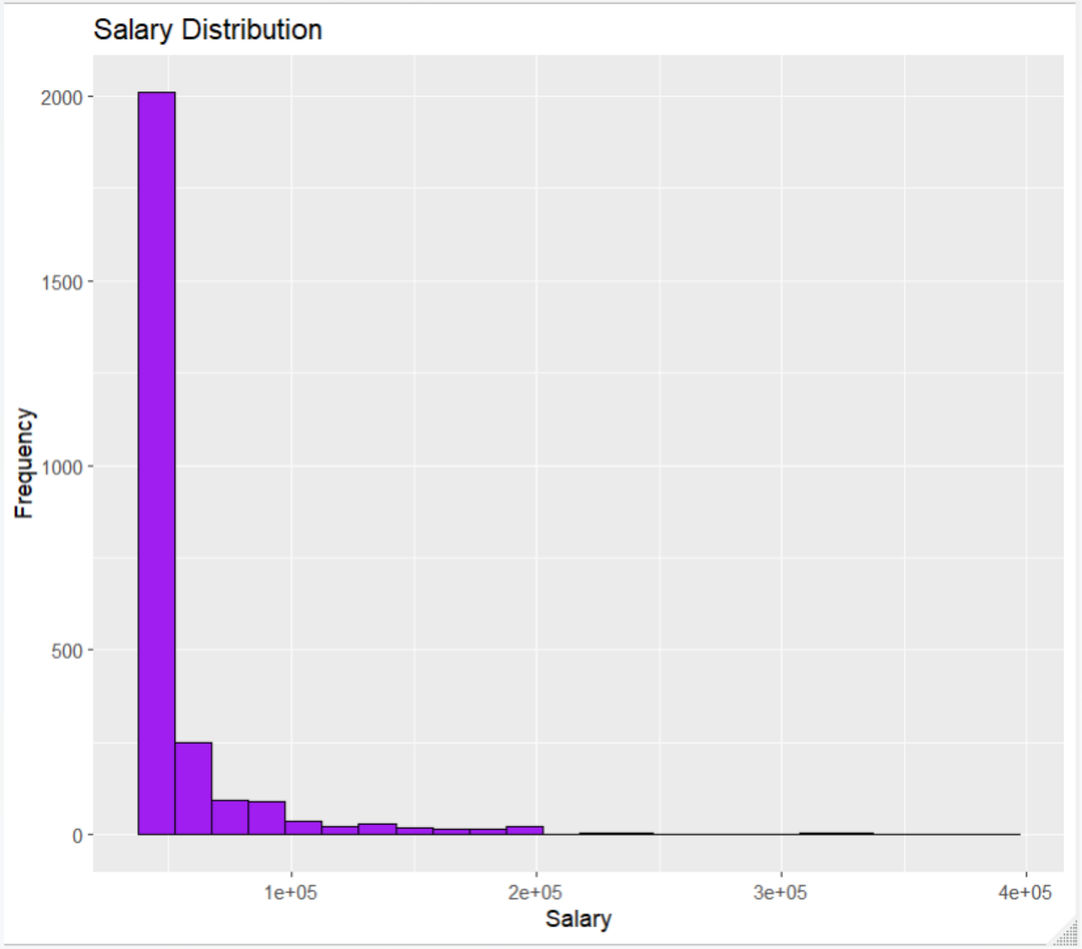
DATA OVERVIEW

Summary Statistics

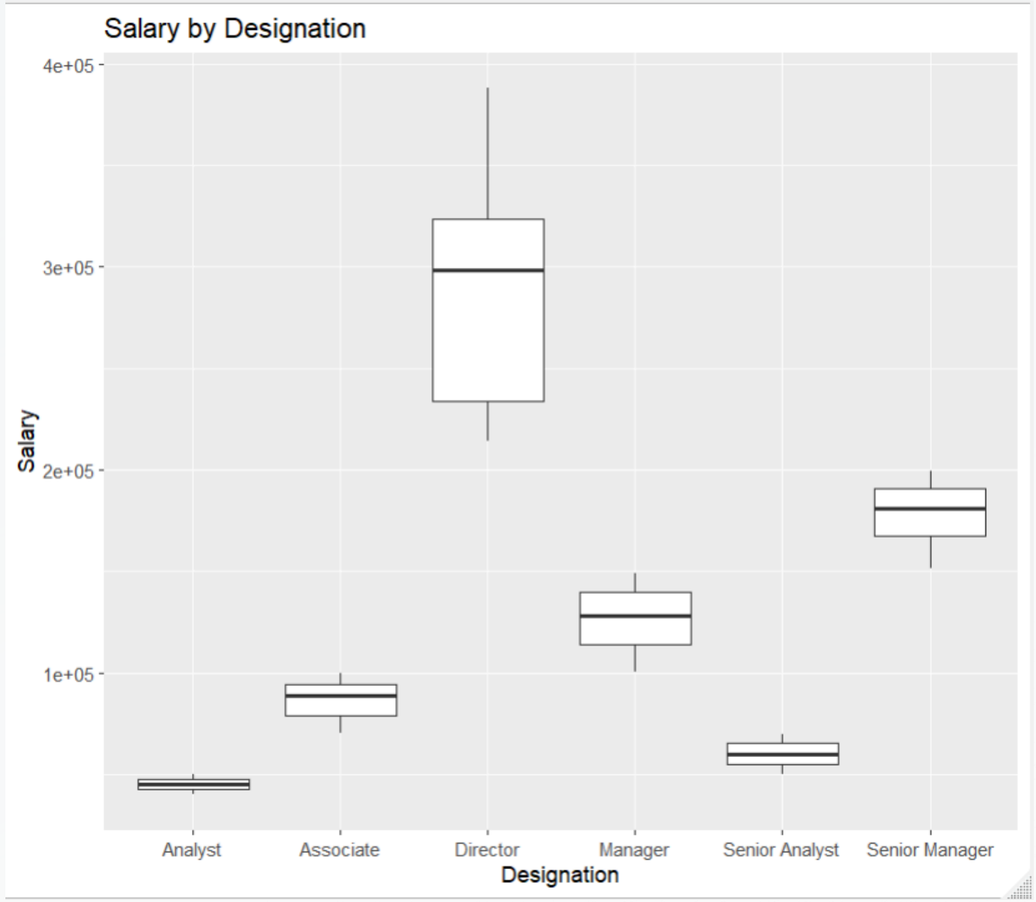
- Minimum Salary: 40,001
- 1st Quartile Salary: 43,418
- Median Salary: 46,780
- Mean Salary: 58,125
- 3rd Quartile Salary: 51,379
- Maximum Salary: 388,112

VISUALS AND INSIGHTS

SALARY DISTRIBUTION HISTOGRAM

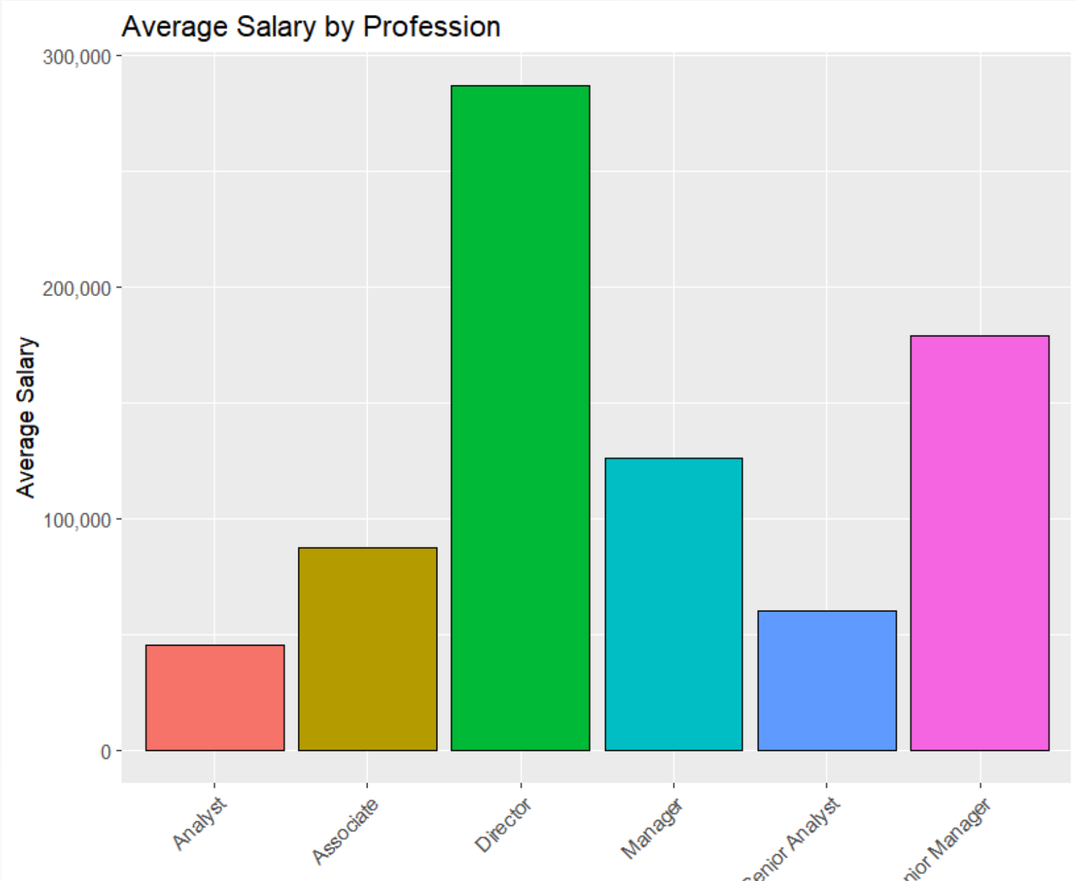


SALARY BY DESIGNATION BOXPLOT

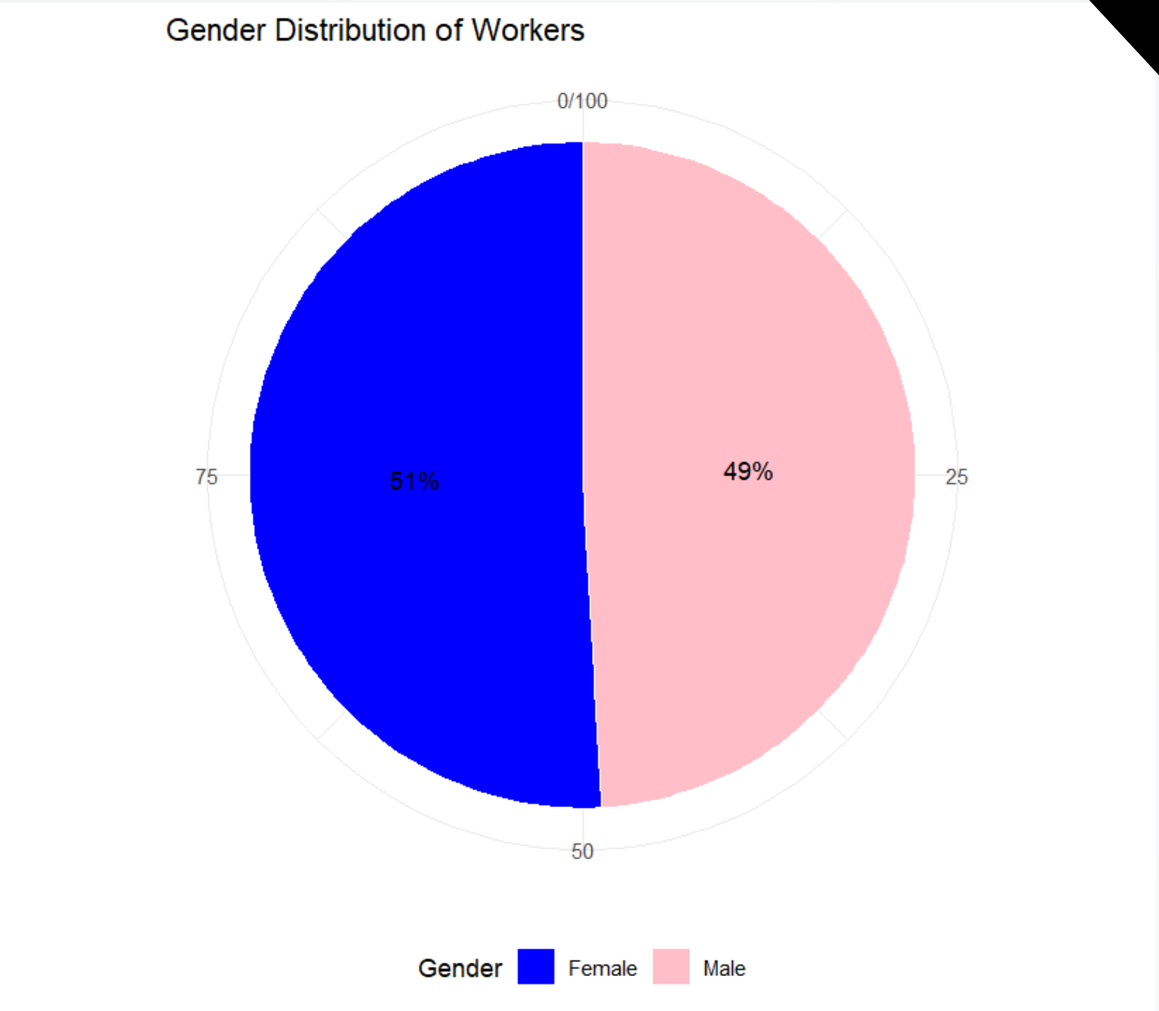


SALARY VS AGE TREND

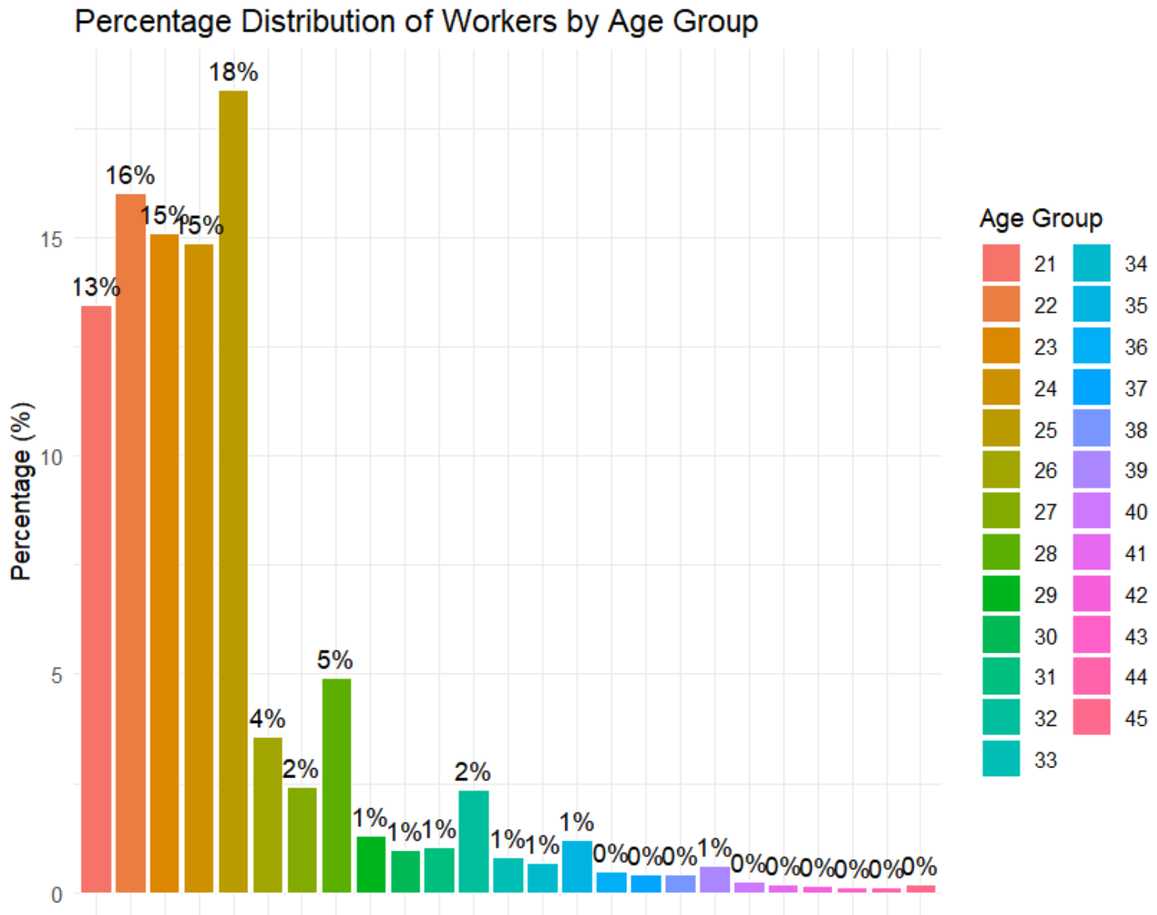
AVERAGE SALARY BY PROFESSION



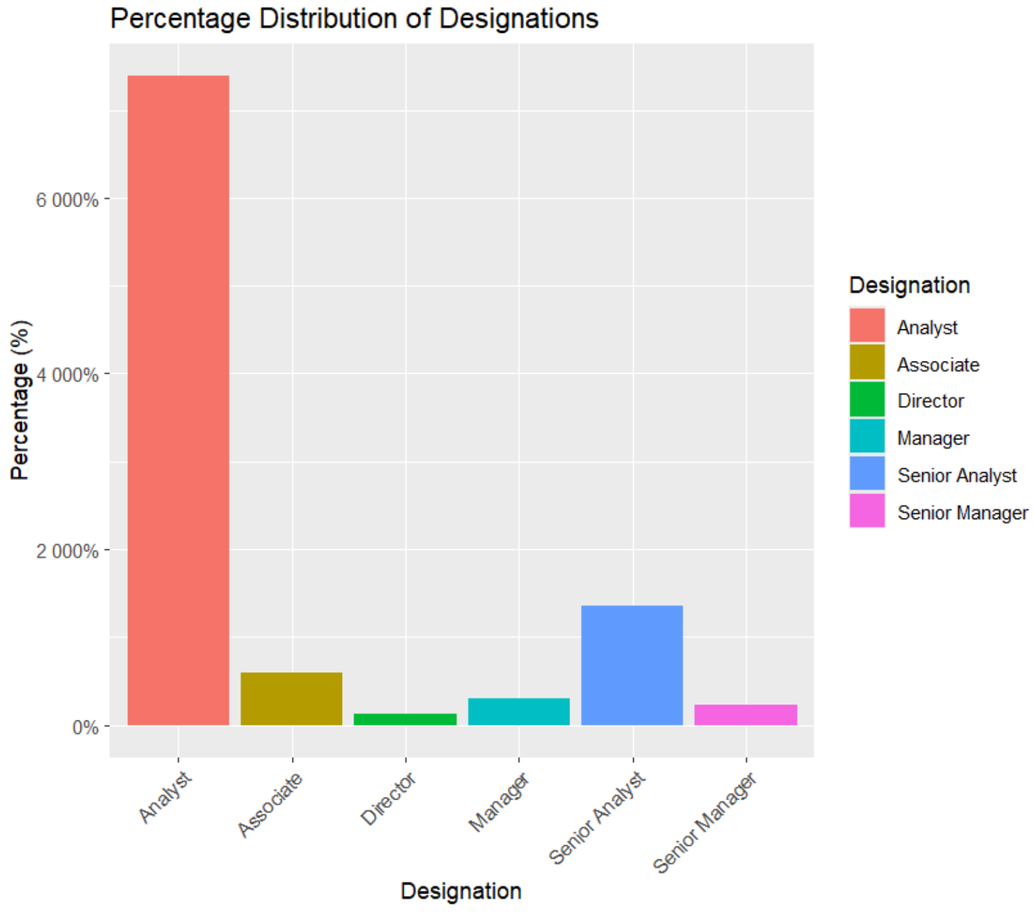
GENDER DISTRIBUTION OF WORKERS



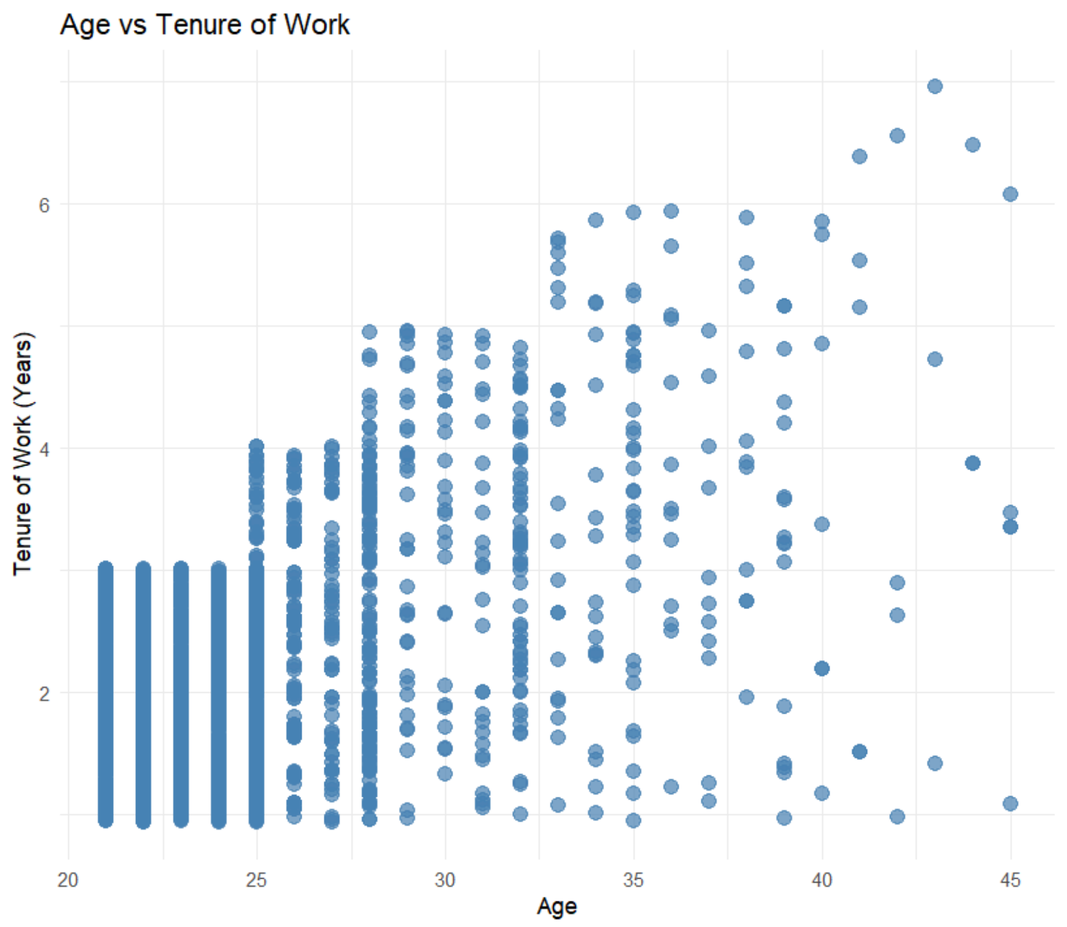
DISTRIBUTION OF WORKERS BY AGE GROUP



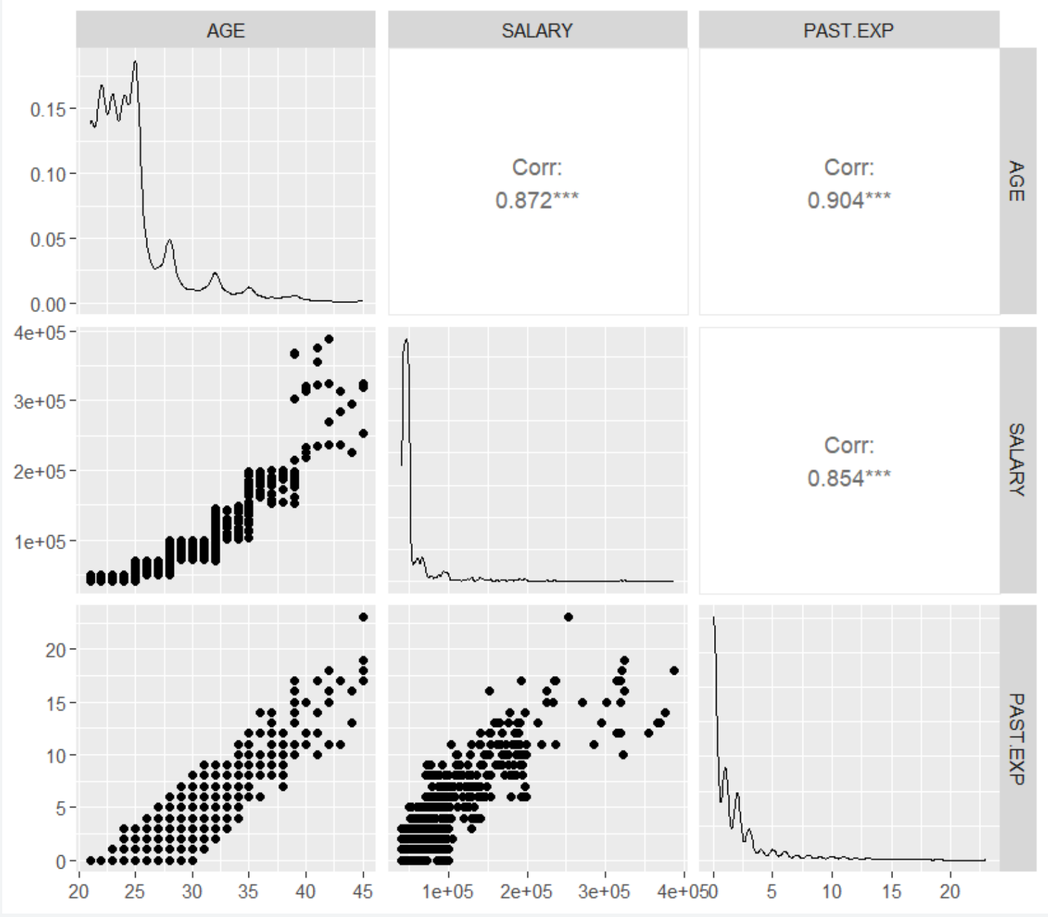
DISTRIBUTION OF WORKERS BY DESIGNATION



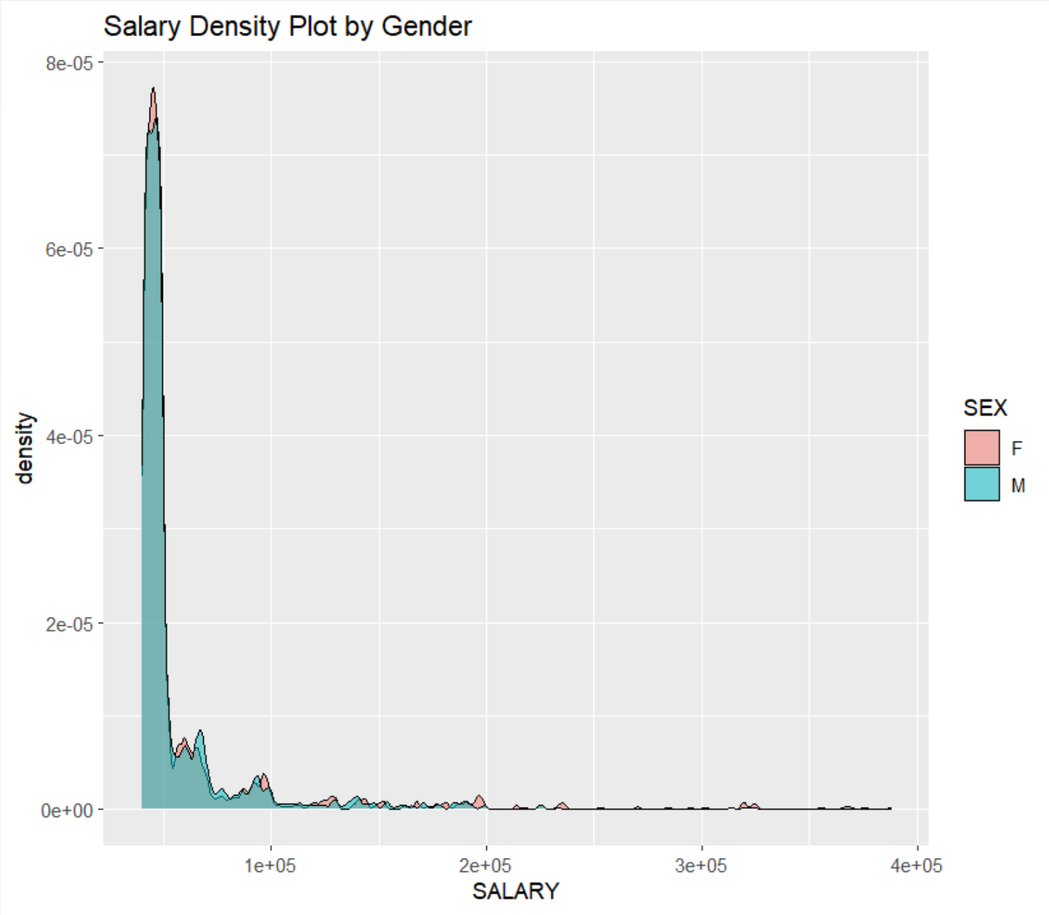
AGE VS TENURE OF WORK



AGE VS SALARY VS PAST EXPERIENCE

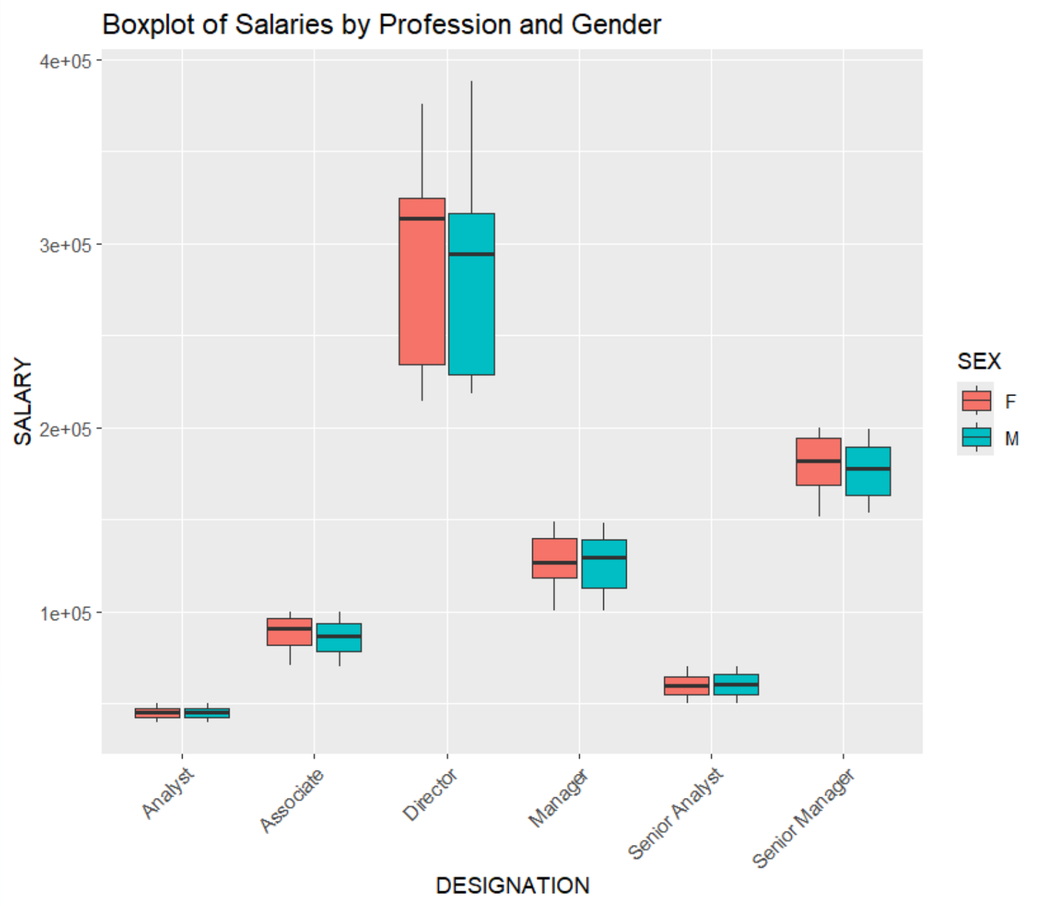


ALARY DENSITY PLOT BY GENDER

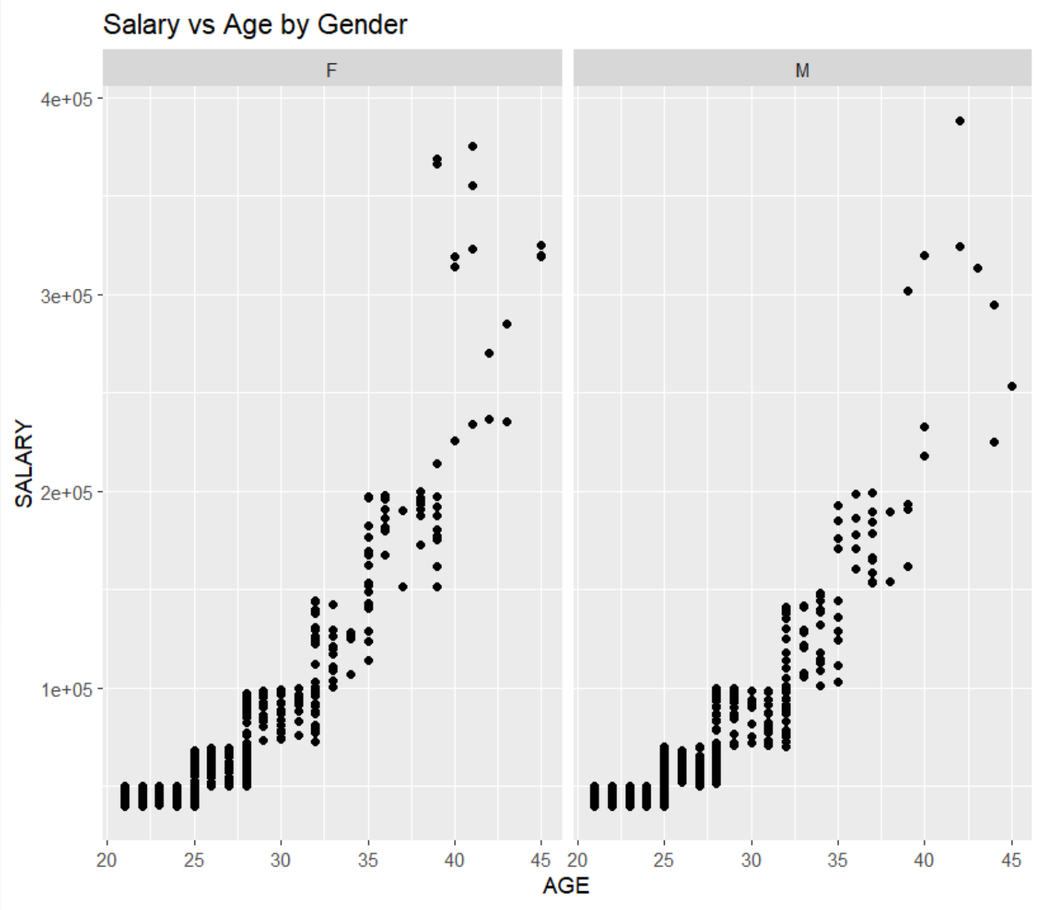


SALARY DISTRIBUTION BY PROFESSION AND GENDER

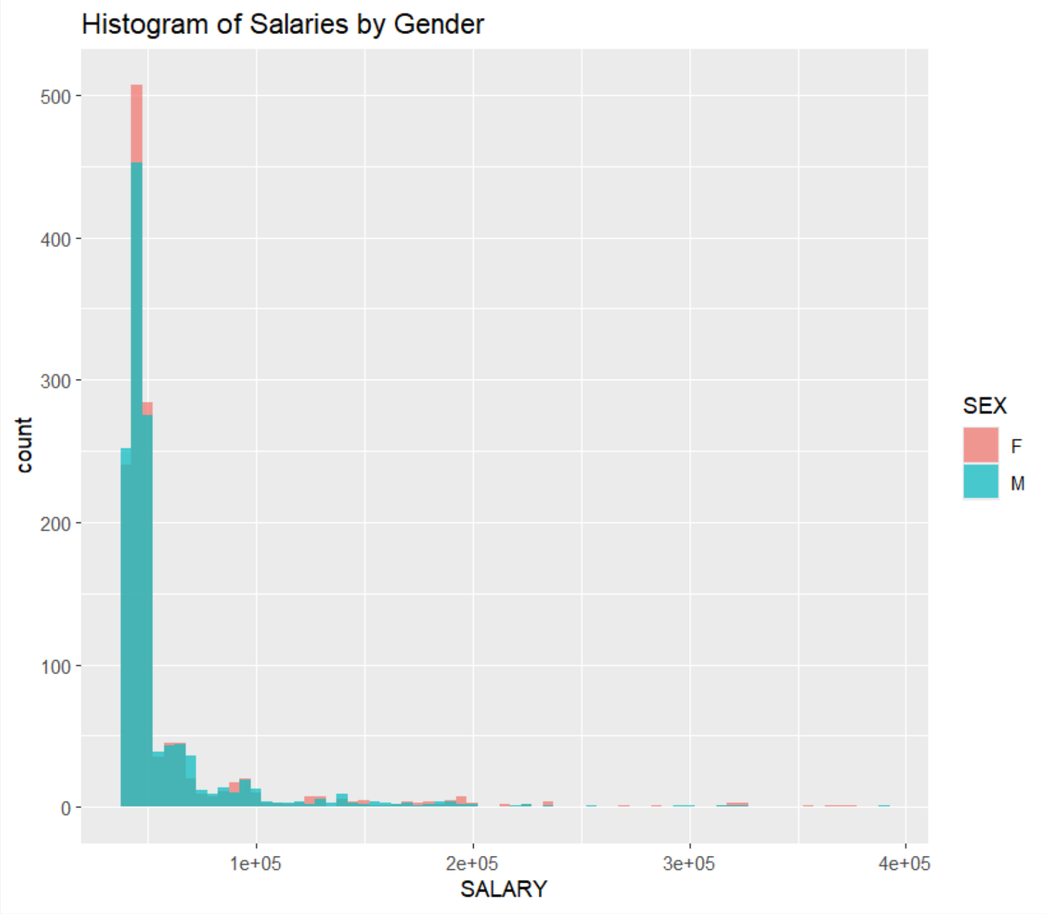
BOXPLOT OF SALARIES BY PROFESSION AND GENDER



SALARY VS AGE BY GENDER

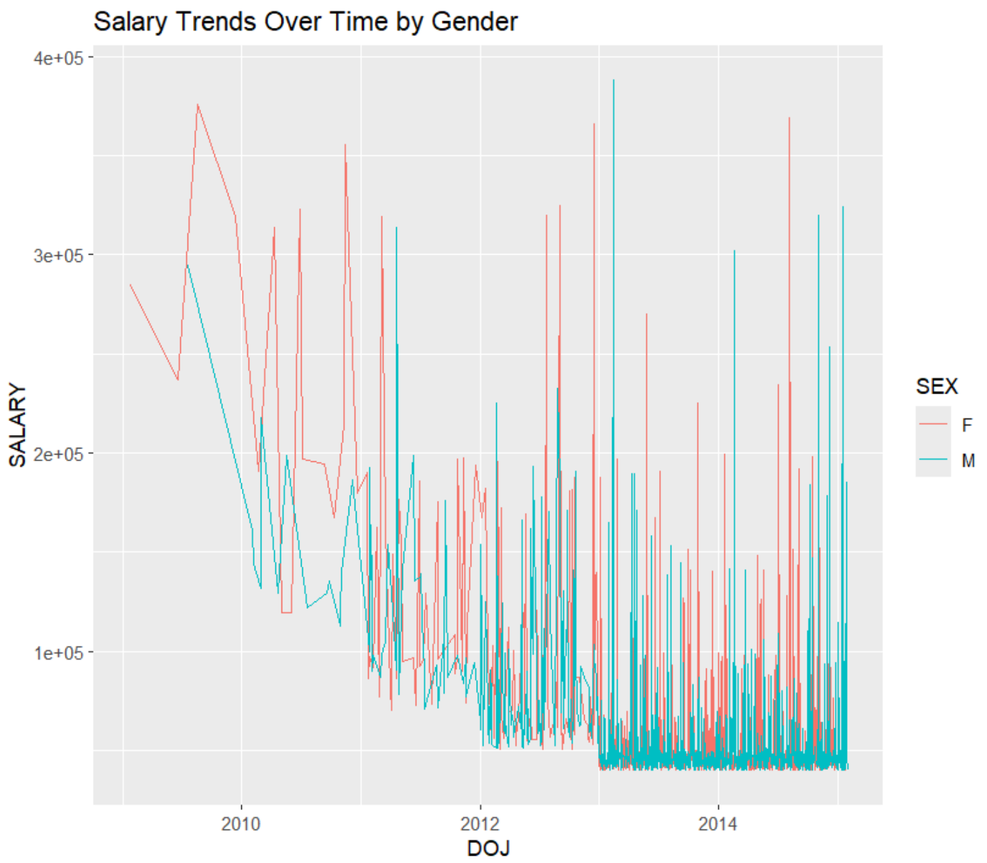


HISTOGRAM OF SALARIES BY GENDER

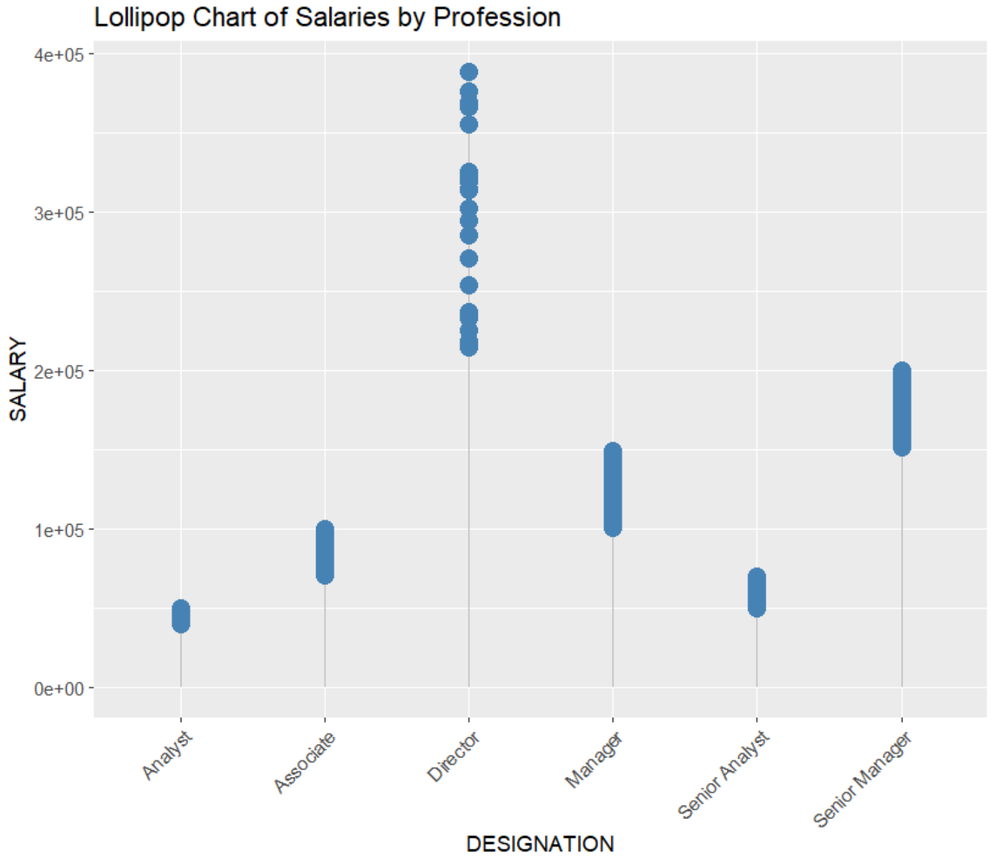


COUNT OF EMPLOYEES

SALARY TRENDS OVER TIME



SALARIES BY PROFESSION



MAXIMUM EARNER DETAILS

```r
# Identify the index of the maximum salary and the corresponding row
max_earner_index <- which.max(salary_data$SALARY)
max_earner <- salary_data[max_earner_index, ]

# Print the details of the maximum earner
print("Details of the Maximum Earner:")
print(max_earner)
```

Filter

	FIRST.NAME	LAST.NAME	SEX	DOJ	CURRENT.DATE	DESIGNATION	AGE	SALARY	UNIT	LEAVES.USED	LEAVES.REMAINING
167	DEREK	ABOULISSAN	M	2013-02-13	2016-01-07	Director	42	388112	Marketing	27	3

MINIMUM EARNER DETAILS

```r
# Identify the index of the maximum salary and the corresponding row
min_earner_index <- which.min(salary_data$SALARY)
min_earner <- salary_data[min_earner_index, ]

# Print the details of the min earner
print("Details of the Min Earner:")

View(min_earner)
```

EDA analysis.R* × | min_earner ×

Filter

	FIRST.NAME	LAST.NAME	SEX	DOJ	CURRENT.DATE	DESIGNATION	AGE	SALARY	UNIT	LEAVES.USED	LEAVES.REMAINING	RATINGS	PAST.EXP
1461	BURL	ALEXIS	M	2014-09-18	2016-01-07	Analyst	22	40001	Operations	22	8	2	0

Gender Distribution

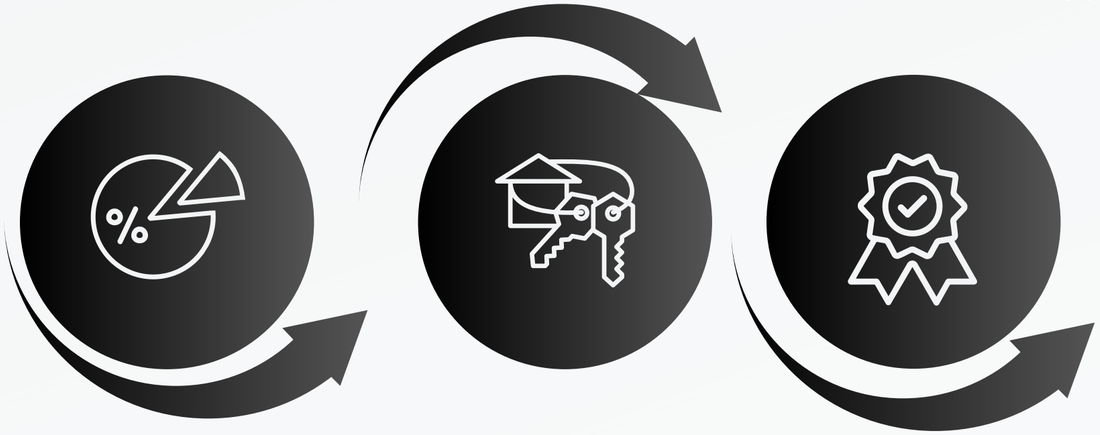
- Higher percentage of male workers compared to female workers.
- Males generally earn higher salaries than females, revealing a gender pay gap.

Average Salary by Profession

- Higher average salaries in professions like Data Scientist and Software Engineer.
- Notable compensation disparities across job titles.

Tenure of Work

- Longer tenure generally associated with higher salaries.
- Experience and longevity are rewarded with higher compensation.



Salary Distribution

- Significant variance in salaries across different designations.
- Majority of salaries clustered in the lower range with a few high-range outliers.

Salary vs. Age

- Positive correlation between age and salary.
- Salary gap between genders persists across age group

Salary by Designation and Gender

- Gender pay gaps within the same designations.
- Males tend to have higher median salaries across most designations.

CONCLUSION

The exploratory data analysis revealed significant insights, including income inequality, gender pay gaps, and the positive correlation between age and salary. These findings highlight key areas for potential improvement and inform strategies for addressing disparities.

THANKYOU !

WWW.LINKEDIN.COM/IN/ALISHBA-RIZWAN--
ALISHBARIZWANAKBARMIRZA@GMAIL.COM