# Assignment 3

Tuan Doğukan TEKEŞ

2025-10-20

```
# Installing necessary package(s)
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```
install.packages("ggmosaic")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```
install.packages("treemapify")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```
installed.packages("scales")
```

```
##      Package LibPath Version Priority Depends Imports LinkingTo Suggests
##      Enhances License License_is_FOSS License_restricts_use OS_type Archs
##      MD5sum NeedsCompilation Built
```

```
library(scales)
library(ggplot2)
library(ggmosaic)
library(treemapify)
```

## Dataset

The `salary` dataset contains information about the average annual salaries of data professionals across different roles and experience levels. It includes numerical data on salaries (`mean_salary`) in Rupees as well as categorical variables describing the job role (`position`) and experience level (`experience`). The experience levels are ordered from Junior → Intermediate → Senior → Executive, reflecting increasing levels of expertise and responsibility. This dataset can be used to explore which positions and experience levels contribute most to overall salary distribution in the data science field.

Source: Kaggle – Data Science Fields Salary Categorization (accessed October 2025).

To load the dataset into your environment, just run the code below:

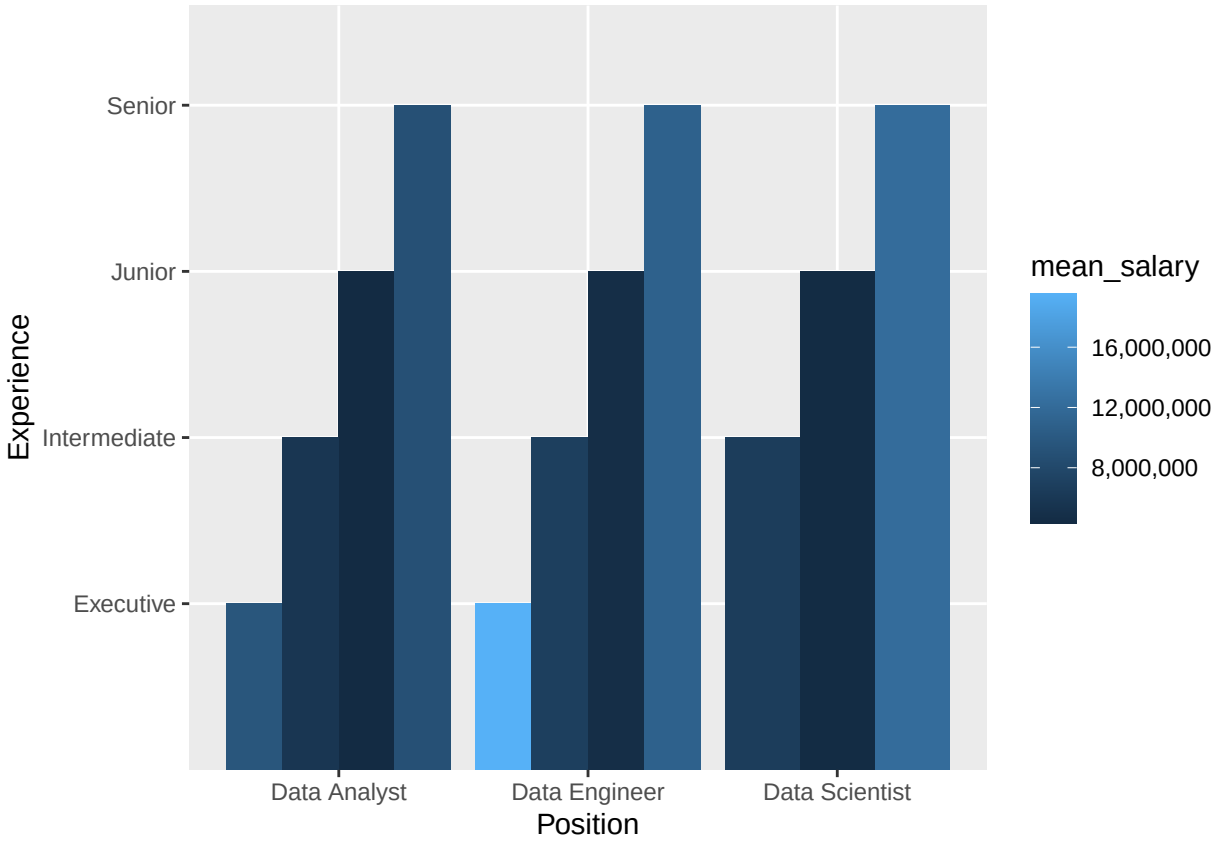
```
salary <- data.frame(
  position    = c("Data Analyst", "Data Analyst", "Data Analyst", "Data Analyst",
                  "Data Engineer", "Data Engineer", "Data Engineer", "Data Engineer",
                  "Data Scientist", "Data Scientist", "Data Scientist"),
  experience  = c("Junior", "Executive", "Intermediate", "Senior",
                  "Junior", "Executive", "Intermediate", "Senior",
```

```
                  "Junior", "Intermediate", "Senior"),
  mean_salary = c(4293623, 9548340, 5705070, 8905628,
                  4689309, 19534312, 6841836, 10903873,
                  4402653, 6527813, 12171827))
```

## Drawing a plot for proportion

1. Draw a plot showing the proportion of total mean salaries by experience and position. The plot should clearly display the hierarchical contribution of each job title within its experience level.

```
ggplot(salary, aes(fill = mean_salary,
                       y    = experience,
                       x    = position)) +
  geom_bar(position = "dodge",
           stat     = "identity")+
  labs ( x = "Position" ,
         y = "Experience") +
  scale_fill_continuous(labels = label_number(big.mark = ",", accuracy = 1))
```




2. Interpret the plot (30 pts). The dataset shows a clear positive relationship between experience and mean salary across all positions as professionals move from Junior to Executive levels, their average pay increases substantially. However, the rate of increase and the absolute salary levels differ notably between roles.