**XSTK BTL**

**Import data**

**Load file**

**setwd("c:/Users/pc/Desktop/Data")**

**export file from csv to rda**

**gh <- read.csv("IT\_Salary\_2020.csv", header = TRUE)**

**save(gh, file = "IT\_salary\_2020.rda")**

**load("c:/Users/pc/Desktop/Data/IT\_salary\_2020.rda")**

**Make separate data in use**

**main\_data <- gh[,c(2,3,5,6,8,9,11,19)]**

**Data Cleaning**

**Change name**

**main\_data\_cleaned = data.frame(Age = main\_data$Age, Gender = main\_data$Gender, Position = main\_data$Position, Experience = main\_data$Total.years.of.experience, Level = main\_data$Seniority.level, Main\_language = main\_data$Your.main.technology...programming.language, Salary = main\_data$Yearly.brutto.salary..without.bonus.and.stocks..in.EUR, Size = main\_data$Company.size)**

**Clear data**

**main\_data\_cleaned <- main\_data\_cleaned[!is.na(main\_data\_cleaned$Age),]**

**main\_data\_cleaned <- main\_data\_cleaned [!is.na(main\_data\_cleaned$Gender),]**

**main\_data\_cleaned <- main\_data\_cleaned [!is.na(main\_data\_cleaned$Position),]**

**main\_data\_cleaned <- main\_data\_cleaned [!is.na(main\_data\_cleaned$Experience),]**

**main\_data\_cleaned <- main\_data\_cleaned [!is.na(main\_data\_cleaned$Level),]**

**main\_data\_cleaned <- main\_data\_cleaned [!is.na(main\_data\_cleaned$Main\_language),]**

**main\_data\_cleaned <- main\_data\_cleaned [!is.na(main\_data\_cleaned$Salary),]**

**main\_data\_cleaned <- main\_data\_cleaned [!is.na(main\_data\_cleaned$Size),]**

**main\_data\_cleaned <- main\_data\_cleaned [!(main\_data\_cleaned$Age == ""),]**

**main\_data\_cleaned <- main\_data\_cleaned [!(main\_data\_cleaned$Gender == ""),]**

**main\_data\_cleaned <- main\_data\_cleaned [!(main\_data\_cleaned$Position == ""),]**

**main\_data\_cleaned <- main\_data\_cleaned [!(main\_data\_cleaned$Experience == ""),]**

**main\_data\_cleaned <- main\_data\_cleaned [!(main\_data\_cleaned$Level == ""),]**

**main\_data\_cleaned <- main\_data\_cleaned [!(main\_data\_cleaned$Main\_language == ""),]**

**main\_data\_cleaned <- main\_data\_cleaned [!(main\_data\_cleaned$Salary == ""),]**

**main\_data\_cleaned <- main\_data\_cleaned [!(main\_data\_cleaned$Size == ""),]**

**combine all command above to clean data**

**main\_data\_cleaned = na.omit(main\_data\_cleaned)**

**Check for missing data**

**sum(is.na(main\_data\_cleaned))**

**Make .csv file**

**write.csv(main\_data\_cleaned, file = "main\_data\_cleaned.csv")**

**summary data**

**summary(main\_data\_cleaned)**