-Good afternoon! Thank you so much for hiring us to explore the Attrition study for Frito Lay.

-Our objective for this study is to do

EDA to understand main factors for Attrition problem

Study Job Role Trends

Build Models to predict Salary and Attrition

Validation of Models for Salary and Attrition

Create a ShinyApp for data visualization

-I’m a mathematician, I love data and I could consider myself as a Data Cruncher and a Hype Crew

- First, by exploring the data, we discover there is no missing data. There are 870 observations with 36 variables.

- The rate in Attrition is 16% or we can say, around 16% employees left jobs.

- The salary range from 1081 to less than 20000

* The salary less than 2000 USD is considered as in lower class.
* The salary from 2000 USD to 4000 USD is considered as in lower middle class and from 4000 USD to 6000 USD is considered as in middle class and from 6000 USD to 10000 USD is considered as in upper middle class.
* The salary from 10000 USD to 16000 USD is considered as in the lower high class and from 16000 USD to 20000 USD is considered as in the high class.
* Then employees with lower Monthly Income will have more chance to leave the current jobs.

- Sales Representatives have the highest rates in Attrition.

- Manufacturing Directors and Reseach Directors have the lowest rates in Attrition.

There are 4 similar variables HourlyRate, DailyRate, MonthlyRate, MonthlyIncome. I will see the relationship of these 3 first variables with MonthlyIncome and these 4 variables with Attrition.

* Weak relationship between HourlyRate/DailyRate/MonthlyRate.
* No meaningful relationship between MonthlyIncome with HourlyRate/DailyRate/MonthlyRate.
* No meaningful relationship between Attrition with HourlyRate/DailyRate/MonthlyRate.

-Later I will use t-test analysis

Attrition rate : At α-level of significance = 0.05, if p-value<0.05 then we reject the null hypothesis (There are some evidence to suggest that the mean difference in the selected variable for two groups Yes-No of Attrition variable) and if p-value>0.05 then we fail to reject the null hypothesis (There are not enough evidence to suggest that the mean difference in the selected variable for two groups Yes-No of Attrition variable).

* By t-test analysis, we will have some ideas to build our models later.
* Similarly I use t test for Gender variable on Job role.
* Numerical Then by some analyses, we can see that Monthly Income have a strong correlation with Job Level. Month Income, Total Working Years and Job Level are correlated strongly. These correlations are logical.
* Automated:
* Then after running this code, I have 14 important variables as follows: Age, Attrition, BusinessTravel, Department, Education, JobLevel, JobRole, NumCompaniesWorked, TotalWorkingYears, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager.

(Income)

Then I will use important variable to build a model later: Age, Department, EnvironmentSatisfaction, JobInvolvement, JobLevel, JobRole, JobSatisfaction, MaritalStatus, MonthlyIncome, NumCompaniesWorked, OverTime, StockOptionLevel, TotalWorkingYears, WorkLifeBalance, YearsAtCompany, YearsInCurrentRole, YearsWithCurrManager.

(Attrition)