## Company introduction:

Idea Spice is a boutique design and marketing firm with offices in the UAE and India. We have contributed to the growth and image of a number of start-ups, SMEs and government organisations.Over 90% of businesses fail in the first 3 years of operation. We realized that having a clear USP or a differentiation tilted the balance in favor of success. We chose to be in the business of increasing the chance of our client's success. We honed the science of finding an empty space in their consumer's mind and connecting with them emotionally. We then translate the brand across all the customer touch points to drive home the business differentiation and create engagement.

## Theme:
Data science

## Project Task 2: Feature Engineering

## Background information

Feature engineering is the process of selecting, manipulating, and transforming raw data into features that can be used in supervised learning. In order to make machine learning work well on new tasks, it might be necessary to design and train better features. You may perform Tweaking features, converting character values into numerical values using appropriate methods**.**

## Dataset Link -
https://drive.google.com/drive/folders/1fmCqKr6DNqy8g3tvlFjMpXPix6f-sLAi?usp=sharing

## Dataset Description -

The columns/features in the given dataset are as follows:

- **EmployeeId:** Company id of an individual employee
- **Age:** Age of an employe (in years)

- **Gender:** Gender of the employee
- **MaritalStatus:** Marital Status of the Employee( Married/ Single/ Divorced etc)
- **Turnover:** Retention rate of an employee
- **Traveling:** Whether traveling is required in job role of an employee
- **Vertical:** In which department employee is working
- **Qualifications:** How many qualifications does an employee have
- **EducationField:** Qualifications are in which field (Medical?marketing/technical etc)
- **EmployeSatisfaction:** How much employee is satisfied in his/her role (1 minimum and 5 maximum)
- **JobEngagement:** How much engagement is required for the role
- **JobLevel:** Level of the job( 1 being entry level and 5 means top management)
- **JobSatisfaction:** How much employee is satisfied with his job role
- **Role:** What is the employee role in the organization (Technician/Manager/HR etc)
- **DailyBilling:** Daily wages of the employees
- **HourBilling:** Hours wages of the employees
- **MonthlyBilling:** Monthly wages of the employee
- **MonthlyRate:** Total salary paid to the employee
- **Work Experience:** Total experience of the employee
- **OverTime:** Whether an employee have done overtime or not
- **PercentSalaryHike:** Last increment
- **Last Rating:** Last financial year performance
- **RelationshipSatisfaction:** Hours wages of the employees
- **Hours:** Working hours of an Employee
- **StockOptionLevel:**
- **TrainingTimesLastYear:** Hours wages of the employees
- **Work&Life:** balance between work and life
- **YearsAtCompany:** Years employee is working with the current organization
- **YearsInCurrentRole:** Years employee working at current designation
- **YearsSinceLastPromotion:** When was last promoted
- **YearsWithCurrentManager:** How long employee has been working with Current manager
- **DistanceFromHome:** Distance between office and employee home

**Task Explanation :**

- Perform outlier analysis on numerical features. Use any method to get rid of outliers
- Look for missing values and duplicate values and find a way to clean it.
- Evaluate the target variable, find patterns and insights from data using visualizations.
- Nature of employees who stay, nature of employees who leave

**Resources -**
Outlier detection:
https://www.freecodecamp.org/news/how-to-detect-outliers-in-machine-learning/


**Submission Format -** Share the file with all the details and insights as **yournametask2.ipynb**
format.