**Background and Context**

You are a Data Scientist for a tourism company named "Visit with us". The Policy Maker of the company wants to enable and establish a viable business model to expand the customer base.

A viable business model is a central concept that helps you to understand the existing ways of doing the business and how to change the ways for the benefit of the tourism sector.

One of the ways to expand the customer base is to introduce a new offering of packages.

Currently, there are 5 types of packages the company is offering - Basic, Standard, Deluxe, Super Deluxe, King. Looking at the data of the last year, we observed that 18% of the customers purchased the packages.

However, the marketing cost was quite high because customers were contacted at random without looking at the available information.

The company is now planning to launch a new product i.e. Wellness Tourism Package. Wellness Tourism is defined as Travel that allows the traveler to maintain, enhance or kick-start a healthy lifestyle, and support or increase one's sense of well-being.

However, this time company wants to harness the available data of existing and potential customers to make the marketing expenditure more efficient.

You as a Data Scientist at "Visit with us" travel company has to analyze the customers' data and information to provide recommendations to the Policy Maker and Marketing Team and also build a model to predict the potential customer who is going to purchase the newly introduced travel package.

**Objective**

To predict which customer is more likely to purchase the newly introduced travel package.

**Data Dictionary**

**Customer details:**

1. CustomerID: Unique customer ID
2. ProdTaken: Product taken flag
3. Age: Age of customer
4. TypeofContact: How customer was contacted (Company Invited or Self Inquiry)
5. CityTier: City tier
6. Occupation: Occupation of customer
7. Gender: Gender of customer
8. NumberOfPersonVisited: Total number of person came with customer
9. PreferredPropertyStar: Preferred hotel property rating by customer
10. MaritalStatus: Marital status of customer
11. NumberOfTrips: Average number of the trip in a year by customer
12. Passport: The customer has passport or not
13. OwnCar: Customers owns a car flag
14. NumberOfChildrenVisited: Total number of children visit with customer
15. Designation: Designation of the customer in the current organization
16. MonthlyIncome: Gross monthly income of the customer

**Customer interaction data:**

1. PitchSatisfactionScore: Sales pitch satisfactory score
2. ProductPitched: Product pitched by a salesperson
3. NumberOfFollowups: Total number of follow up has been done by sales person after sales pitch
4. DurationOfPitch: Duration of the pitch by a salesman to customer

**Submission Guidelines :**

1. **There are two parts to the submission:**
   1. **A well commented Jupyter notebook [format - .ipynb]**
   2. **A presentation as you would present to the top management/business leaders [format - .ppt /.pptx]**

**Scoring guide (Rubric) - Travel Package Purchase Prediction**

| **Criteria** | **Points** |
| --- | --- |
| **Perform an Exploratory Data Analysis on the data**  - Univariate analysis - Bivariate analysis - Use appropriate visualizations to identify the patterns and insights - Come up with a customer profile (characteristics of a customer) of the different packages - Any other exploratory deep dive  **3 Slides on Presentation** | 7.5 |
| **Illustrate the insights based on EDA**  Key meaningful observations on the relationship between variables  **3 Slides on Presentation** | 5 |
| **Data Pre-processing**  Prepare the data for analysis - Missing value Treatment, Outlier Detection(treat, if needed- why or why not ), Feature Engineering, Prepare data for modeling  **2 Slides on Presentation** | 7.5 |
| **Model building - Bagging**  - Build bagging classifier, random forest and decision tree.  **1 Slide on Presentation** | 6 |
| **Model performance evaluation and improvement**  - Comment on which metric is right for model performance evaluation and why? - Comment on model performance - Can model performance be improved? check and comment  **2 Slides on Presentation** | 9 |
| **Model building - Boosting**  - Build Adaboost, gradient boost, xgboost and stacking classifier  **1 Slide on Presentation** | 8 |
| **Model performance evaluation and improvement**  - Comment on which metric is right for model performance evaluation and why? - Comment on model performance - Can model performance be improved? check and comment  **2 Slides on Presentation** | 12 |
| **Actionable Insights & Recommendations**  - Compare models - Business recommendations and insights  **3 Slides on Presentation with Actual Business Insights and Corresponding Recommendations to be Actioned** | 5 |
| Points | 60 |