Travel Package Purchase Prediction - Problem Statement

**Description**

**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 have 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: Whether the customer has purchased a package or not (0: No, 1: Yes)
3. Age: Age of customer
4. TypeofContact: How customer was contacted (Company Invited or Self Inquiry)
5. CityTier: City tier depends on the development of a city, population, facilities, and living standards. The categories are ordered i.e. Tier 1 > Tier 2 > Tier 3
6. Occupation: Occupation of customer
7. Gender: Gender of customer
8. NumberOfPersonVisiting: Total number of persons planning to take the trip with the customer
9. PreferredPropertyStar: Preferred hotel property rating by customer
10. MaritalStatus: Marital status of customer
11. NumberOfTrips: Average number of trips in a year by customer
12. Passport: The customer has a passport or not (0: No, 1: Yes)
13. OwnCar: Whether the customers own a car or not (0: No, 1: Yes)
14. NumberOfChildrenVisiting: Total number of children with age less than 5 planning to take the trip with the 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 satisfaction score
2. ProductPitched: Product pitched by the salesperson
3. NumberOfFollowups: Total number of follow-ups has been done by the salesperson after the sales pitch
4. DurationOfPitch: Duration of the pitch by a salesperson to the customer

**Note:**

Please note XGBoost can take a significantly longer time to run, so if you have time complexity issues then you can avoid tuning XGBoost. No marks will be deducted if XGBoost tuning is not attempted.

**Best Practices for Notebook :**

* The notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
* The notebook should be run from start to finish in a sequential manner before submission.
* It is preferable to remove all warnings and errors before submission.
* The notebook should be submitted as an HTML file (.html) and as a notebook file (.ipynb)

**Submission Guidelines :**

1. There are two parts to the submission:
   1. A well commented Jupyter notebook [format - .ipynb]
   2. File converted to HTML format
2. Any assignment found copied/ plagiarized with other groups will not be graded and awarded zero marks
3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation
4. Submission will not be evaluated if,
   1. it is submitted post-deadline, or,
   2. more than 2 files are submitted

Happy Learning!!

**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 | 8 |
| **Illustrate the insights based on EDA**  Key meaningful observations on individual variables and the relationship between variables | 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 | 8 |
| **Model building - Bagging**  - Build Bagging classifier, Random Forest, and Decision Tree. - Comment on model performance | 6 |
| **Model performance improvement - Bagging**  - Comment on which metric is right for model performance evaluation and why? - Comment on the model performance after tuning the Decision Tree, Bagging, and Random Forest classifier to improve the model performance. | 9 |
| **Model building - Boosting**  - Build Adaboost, GradientBoost, XGBoost, and Stacking classifiers - Comment on model performance | 6 |
| **Model performance improvement - Boosting**  - Comment on which metric is right for model performance evaluation and why? - Comment on the model performance after tuning the AdaBoost, and Gradient Boosting classifier on the appropriate metric to improve the model performance. \* Please note XGBoost can take a significantly longer time to run, so if you have time complexity issues then you can avoid tuning XGBoost. | 7 |
| **Actionable Insights & Recommendations**  - Compare model performance on various metrics. - Conclude with the key takeaways - What would your advice be to grow the business? | 6 |
| **Notebook - Overall**  - Structure and flow - Well commented code | 5 |
| Points | 60 |