

## High Level Design (HLD)

# Travel Package Purchase Prediction

**Revision No: 1.0**

**Last Date of Revision: 25/05/23**

**Manasi Khillare**

**Naresh K**

## Document Version Control

Date Issued	Version	Description	Author
25/05/23	1.0	First Version of Complete HLD	Manasi Khillare Naresh K

# Contents

<b>Abstract .....</b>	<b>3</b>
<b>1. Introduction.....</b>	<b>4</b>
1.1 Why this High-Level Design Document?	
1.2 Scope	
<b>2. General Description.....</b>	<b>6</b>
2.1 Product Perspective	
2.2 Problem Statement	
2.3 Proposed Solution	
2.4 Further Improvements	
2.5 Technical Requirements	
2.6 Data Requirements	
2.7 Tools Used	
2.8 Constraints	
<b>3. Design Details.....</b>	<b>7</b>
3.1 Process Flow	
3.1.2 Deployment Process	
<b>4. Performance.....</b>	<b>8</b>
4.1 Reusability	
4.2 Application Compatibility	
4.3 Resource Utilization	
4.4 Deployment	
<b>5. Conclusion.....</b>	<b>9</b>

## Abstract

Travel has always been a unique experience to every individual. Expanding the possibilities of these unique experiences has been the goal of every travel company throughout the years. Aiding in travel and tourism, for individuals to explore cultures, traditions and beauty of places they visit, in the most customer - oriented way possible, has been the goal.

Tourism is one of the most rapidly growing global industries and tourism forecasting is becoming an increasingly important activity in planning and managing the industry. Because of high fluctuations of tourism demand, accurate predictions of purchase of travel packages are of high importance for tourism organizations. The goal is to predict whether the customer will purchase the travel or not.

The Machine Learning system design is to predict whether the customer will or will not purchase a particular package. So, to analyze the data several Machine Learning algorithms. A comprehensive prediction is done using Machine Learning models.

# 1 Introduction

## 1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
  - o Security
  - o Reliability
  - o Maintainability
  - o Portability
  - o Reusability
  - o Application compatibility
  - o Resource utilization
  - o Serviceability

## 1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

## 2 General Description

### 2.1 Product Perspective

Travel Package Purchase Prediction is a solution that is able to predict the purchase of the different travel packages according to the provided dataset.

### 2.2 Problem Statement

Tourism is one of the most rapidly growing global industries and tourism forecasting is becoming an increasingly important activity in planning and managing the industry. Because of high fluctuations of tourism demand, accurate predictions of purchase of travel packages are of high importance for tourism organizations. Using the classical machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing, the goal is to predict whether the customer will purchase the travel package or not.

### 2.3 Proposed Solution

We will perform EDA to find the important relation between different attributes and will use a machine-learning algorithm to predict whether the customer will purchase the pitched product or not. The user will be filling the required feature as input and will get results through the web application. The system will then get features which would be passed on for validation and pre-processing. Lastly, the processed information will be passed to a hyperparameter-tuned machine learning model to predict the final outcome.

## 2.4 Data Requirements

The Data requirements completely depend on our problem statement. The excel data consists of 4888 rows. The features it holds are as follows:

- **CustomerID:** Unique customer ID **ProdTaken:** Whether the customer has purchased a package or not (0: No, 1: Yes)
- **Age:** Age of customer **TypeofContact:** How customer was contacted (Company Invited or Self Inquiry)
- **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
- **Occupation:** Occupation of customer
- **Gender:** Gender of customer
- **NumberOfPersonVisiting:** Total number of persons planning to take the trip with the customer
- **PreferredPropertyStar:** Preferred hotel property rating by customer
- **MaritalStatus:** Marital status of customer
- **NumberOfTrips:** Average number of trips in a year by customer
- **Passport:** The customer has a passport or not (0: No, 1: Yes)
- **OwnCar:** Whether the customers own a car or not (0: No, 1: Yes)
- **NumberOfChildrenVisiting:** Total number of children with age less than 5 planning to take the trip with the customer
- **Designation:** Designation of the customer in the current organization
- **MonthlyIncome:** Gross monthly income of the customer
- **Customer interaction data:**
- **PitchSatisfactionScore:** Sales pitch satisfaction score
- **ProductPitched:** Product pitched by the salesperson
- **NumberOfFollowups:** Total number of follow-ups has been done by the salesperson after the sales pitch
- **DurationOfPitch:** Duration of the pitch by a salesperson to the customer

## 2.5 Technical Requirements

The solution is a cloud-based web app hosted on. For accessing this application below are the minimum requirements:

- Good internet connection.
- Web Browser.

For training model, the system requirements are as follows:

- +4 GB RAM preferred High Level Document (HLD) 7
- Operation System: Windows, Linux, Mac
- Pycharm

## 2.6 Tools Used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Flask are used to build the whole model.



- PyCharm and Google Colab is used as IDE.
- For visualization of the plots Matplotlib, and Seaborn are used.
- Flask is used for the deployment of the model.
- Front end Development is done using HTML/CSS, and Bootstrap 4.
- Python is used for backend development.
- Github is used as a version control system.
- The project is deployed using Heroku Platform

## 2.7 Further Improvements

As a future scope, features like data visualizations after model training can be done to understand the customers better.

## 2.8 Constraints

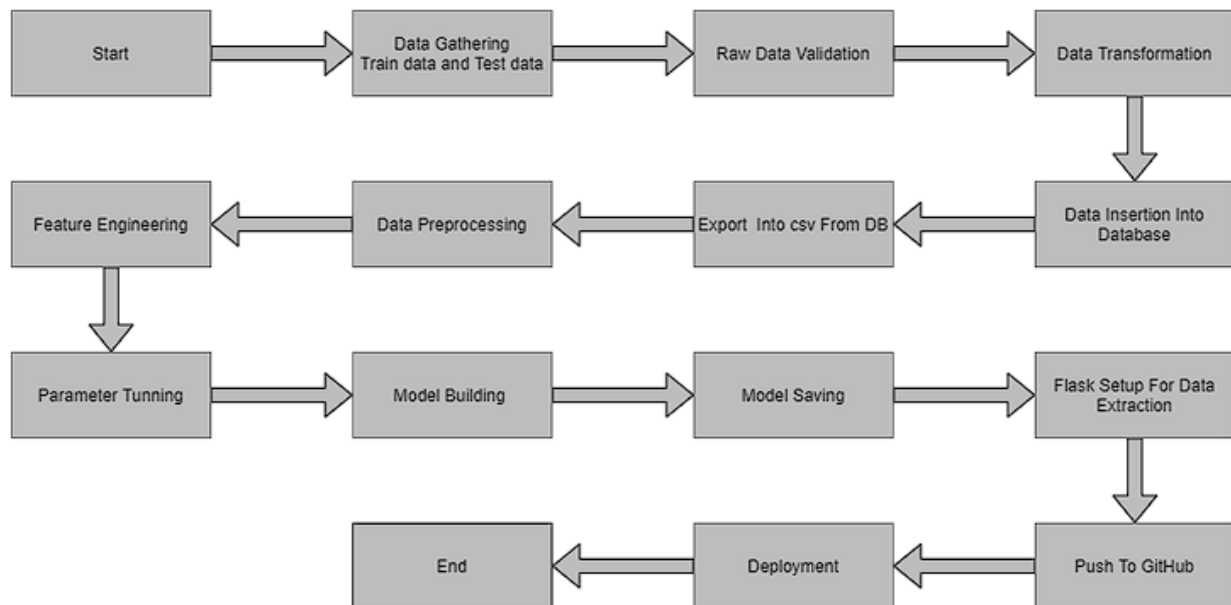
The Travel Package Purchase Prediction system must be user-friendly, and as automated as possible and users should not be required to know any of the workings.



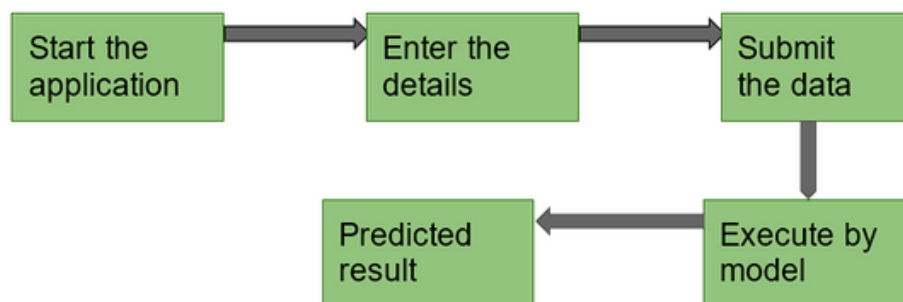
### 3 Design Details

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Flask are used to build the whole model.

#### 3.1 Process Flow



#### 3.2 Deployment Process



## 4 Performance

### 4.1 Reusability

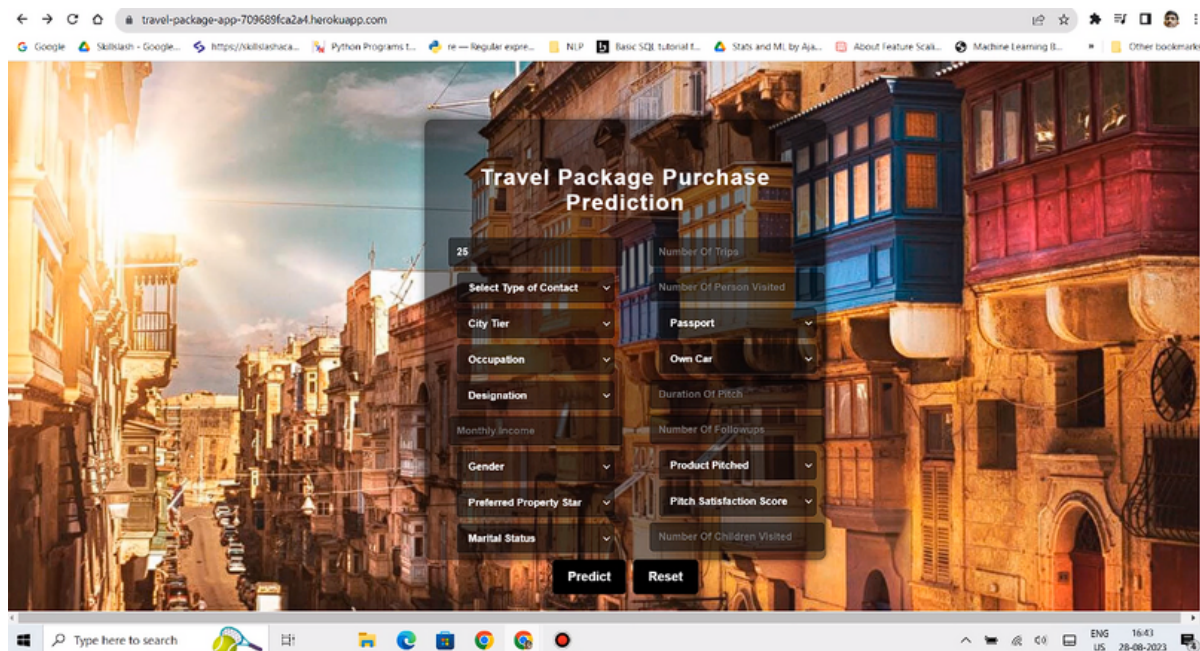
The entire solution will be done in modular fashion and will be API oriented. So, in the case of the scaling the application, the components are completely reusable.

### 4.2 Application Compatibility

The interaction with the application is done through the designed user interface, which the end user can access through any web browser.

### 4.3 Deployment

The application is deployed using Heroku Platform



## 5 Conclusion

Travel data has been cleaned using Python and analyzed which will provide a better key insight for the data and tell you a better story of the raw data.

The same travel data is used to build a prediction model. Which is finally deployed as a web application