Crime Analysis in Power BI

# Revision Number: 1.0

Last date of revision: 15/05/2022

Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
| 15/05/2022 | 1 | Initial HLD — V1.0 | Prateek Jha |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Contents**

[Document Version Control 2](#_TOC_250021)

[Abstract 4](#_TOC_250020)

1. [Introduction 5](#_TOC_250019)
   1. [Why this High-Level Design Document? 5](#_TOC_250018)
   2. [Scope 5](#_TOC_250017)
   3. [Definitions 5](#_TOC_250016)
2. [General Description 6](#_TOC_250015)
   1. [Product Perspective 6](#_TOC_250014)
   2. [Problem statement 6](#_TOC_250013)
   3. [PROPOSED SOLUTION 6](#_TOC_250012)
   4. [FURTHER IMPROVEMENTS 6](#_TOC_250011)
   5. [Technical Requirements 6](#_TOC_250010)
   6. [Data Requirements 7](#_TOC_250009)
   7. [Tools used 8](#_TOC_250008)
      1. Hardware Requirements 8
      2. ROS(Robotic Operating System) 9
   8. [Constraints 9](#_TOC_250007)
   9. [Assumptions 9](#_TOC_250006)
3. [Design Details 10](#_TOC_250005)
   1. [Process Flow 10](#_TOC_250004)
      1. [Model Training and Evaluation 10](#_TOC_250003)
      2. [Deployment Process 11](#_TOC_250002)
   2. [Event log 11](#_TOC_250001)
   3. [Error Handling 11](#_TOC_250000)
   4. Performance 12
   5. Reusability 12
   6. Application Compatibility 12
   7. Resource Utilization 12
   8. Deployment 12
4. Dashboards 13
   1. KPls (Key Performance lndicators) 13
5. Conclusion 14

# Abstract

The airline corporations use complex strategies for flight ticket’s price prediction. The highly complex method makes the customers very difficult to guess the exact or near to exact price for the day of travelling as per their convenience, as the flight fare change dynamically.

Our project is an improved Flight Fare Prediction app which will solve this problem and we will try to make it easier for our customers to predict the ticket fare before travelling with least hassle possible and can book tickets from any airline as per their comfort.

# Introduction

## **Why this High-Level Design Document:**

The main goal of this high level design (HLD) document is to detect all the inconsistencies in the design prior to code implementation and it can be used as a reference manual in case we need to understand the working of any manual for any of the modules.

This HLD will contain the following information needed:

* + - Present all of the design aspects and define them in detail
    - Describe the user interface being implemented
    - Description of user interfaces
    - Description of performance requirements
    - Project design as a whole unit
    - List and describe the non-functional attributes like:
      * + Reliability
        + Maintainability
        + Portability
        + Reusability
        + Application compatibility
        + Resource utilization
        + Serviceability

## Scope

This HLD documentation covers the overall architecture of the system, such as the database usage, application’s flow layer by layer both from UI and backend perspective and technology architecture. This HLD will contain non-technical to slightly technical terms related to the work but we will explain all of this here completely from scratch without any problem.

## Definitions

*Term Description*

|  |  |
| --- | --- |
| *PBI*  *DB*  *ETL*  *CSV* | Microsoft Power BI |
| DB is the short form for database or collection of all the information monitored by this system |
| Used to denote the term Extract Transform and Load |
| Source files containing data to be put in MySQL database |

# General Description

## **Product Perspective**

The Flight Fare Prediction app is a machine learning model which will help customers planning an air travel find the best flight fare with appropriate conditions set as per their requirements. The Power BI report designed for the available dataset is created to give a more detailed and visual interpretation of criminal activities happening all over city of Baton Rouge.

### **Problem statement**

We are trying to give the users a self service BI report that will give the users a brief idea as to help them focus on areas where most offensive activities happening and how to improve and reduce these incidents in any way possible.

It is very difficult to know about the overall idea just by looking at the raw file data in best possible manner but everyone understands data better in a visually appealing way. So we are designing a report which would resolve all those issues in understanding data in raw form as it is generated.

### **PROPOSED SOLUTION**

The solution proposed here is an UGV (Unmanned Ground Vehicle) based Surveillance (Unmanned Ground Vehicle) can be implemented to perform above mention use cases .In first case, if UGV detects any mob(illegal) activities at a particular location it will take photos or video for the evidence and send the police the current location where the mob activities are taking place, further in the second use case, if UGV detects any natural or human made disasters (fire, smoke, etc..) the UGV detects with its sensors and will send details to concerned authorities and lastly in the final use case of UGV, if it finds any medical emergency (accident, etc..)it will take rapid action (call ambulance and alert the nearest hospital )for swift help.

We are proposing a solution where users with no or little background knowledge of machine learning, can simply come to this portal that we’ve opened for our users, enter some details based on the preferences from available list of options can get the most accurate results for fare prediction with very less/no deviation from the original price that we can estimate via manual methods.

The proposed solution here is a Power Bi report which is hosted in Power BI service which will capture all data changes from the source files generated for the same, perform an ETL using python and query editor steps in Power BI itself to clean the data as much we can and using the logic appropriately for different visuals for representing different needs.

### **FURTHER IMPROVEMENTS**

This fare prediction system has certain room for improvement such as addition of a flag feature to understand if the journey is going to be on weekend or weekdays and we can add the same for national or international holidays in the backend side for the model to be able to make more accurate predictions. Since the current model is only trained on national flight data, we need to find more data if we want to make fare predictions for international flights from one source to another destination. If our web app does not perform up to the mark, we need to use another machine learning model to use the existing data for better results where performance is more.

Of course, this report will need to be refreshed with more data so a new file will be needed everytime with updated data which will be processed via python script to be dumped in MySQL DB and then finally to be refreshed in PBI service which is on cloud.

## **Technical Requirements**

* + - Data requirement completely depend on our problem statement.
    - We data in a relational table format containing various columns/features where we will have one or multiple columns which will be used for prediction and there will be only one dependent feature/column which will be used for prediction.
    - We will need some columns as independent columns such as OffenseDate, OffenseTime,

CrimeCategory, Committed, Address related information to incident, zones where incident happened and category of offense done.

* + - The source of all these files will be a CSV file and after collection of entire data, it will dumped into a MySQL DB after some cleaning is done via python script.
    - As far as technical knowledge is concerned, we need to be fundamentally cleared up on basics of some DAX functions in PBI for calculated columns and measures, pandas and python data manipulation using lists
    - Working knowledge of MySQL DB as this is the main database which we will be using for data related activities.

## **Data Requirements**

* + - Data requirement completely depend on our problem statement.
    - We data in a relational table format containing various columns/features where we will have one or multiple columns which will be used for prediction and there will be only one dependent feature/column which will be used for prediction.
    - The source of all these files will be a CSV file and after collection of entire data, it will dumped into a MySQL DB and from there, we will fetch it for cleaning and import in PBI accordingly needed per purpose.

## **Tools used**

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn are used to build the entire backend part for machine learning modelling part.



* + - IDE to be used for development for Python script for ETL purpose is Jupyter notebook or any IDE that support python scripting
    - MySQL DB is used for storage and retrieval from source database
    - GitHub is used as version control system
    - Microsoft Power Bi for April 2022 release is currently used for the development of the report but can be used for changes in any upcoming versions of the release for this product.
* ~~- - - •~~ High Level Design 

## **Constraints**

The report created should be user friendly and should provide the user with correct details for which user is using the platform for with best accurate data to be shown in visual format.

**- 1**High Level Design (HLD) 

# Design Details

## **Process Flow**

For our Crime analysis report to work properly, we will use a python script to fetch data from csv and rest of steps and explained in detail below

Proposed methodology

