

House Price Prediction

Project Charter

By: [Cc]

# Project Charter Approvals

# Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Name | Comments | Version |
| <Date> | <Name> | <Comments> | 1.0 |

1. About the Project Charter

# Project Authorization

The primary purpose of the project charter is for the allocation of resources and authorization to commence work. A project’s charter is typically evaluated against others and only select projects are authorized to proceed. Therefore, the charter should properly highlight the benefits of the project as well as potential risks.

# ***Audience***

The intended audience of the project charter is the decision makers who will allocate resources to the project. However, many others in the organization typically see the project charter as well including project team members, direct and indirect stakeholders, and staff in related groups and departments. The author should keep these additional viewers in mind when writing the charter.

1. Project Overview

# Executive Summary

<The Executive Summary should provide a high level summary of the project in one page or less with text or visuals as needed>

When we have house attribute data and sales price data, if we can relate to each other, this is very important for real estate websites and house buyers.

In addition, this model can help the recommendation system function more usefully: Knowing the predicted price is very important for house recommendations for price-sensitive groups.

1. Project Objectives and Expected Benefits

# Objectives – <Enter the Objectives>

<Describe the objective in more detail>

Everybody needs a roof over their heads. It can be a house, a villa, or a flat. Everybody, at some point in life, faces a choice whether to buy a house, and if so, which one. And why are they so expensive?

The topic of real estate is not only the topic you just have to deal with. It can also be very interesting. There are plenty of TV Shows, for instance, [*Property Brothers*](https://en.wikipedia.org/wiki/Property_Brothers), of which plot is based on examples of people buying and renovating houses. This particular one is the most famous in the world and has been running already for almost a decade. For many people houses are also an investment that generates profits.

Regardless of motives of buying and selling real estate, both sides agree on a price. It is always good to know **how much** a house is worth, what is the expected transaction price. Furthermore, it may be even more important **why** is the price like that, what has an impact on it.

In this work, we want to find an answer to both questions with a stronger emphasis on the latter. This paper intends to be a comprehensive use case of how to deal with a regression problem for Data Scientists. Let us start with a couple of questions that allow to define and understand problems regarding house pricing.

* *The seller does not know how to increase the value of the apartment so that the investment outlay is lower than the added value (e.g. building a pool may increase the price and renovating the bathroom is not worth it).*
* *The seller does not know how much to sell the apartment for (he makes an offer on the portal and does not know if the price is adequate).*
* *The buyer does not know how much the apartment is worth (as above, whether the price is adequate).*
* *Commercial problem: auction services may be interested in tools to support sellers and buyers (to highlight the sections in the offers that most affect the price).*

These are just some of the questions we can ask. As a definition of our problem, we set the property valuation, and through explanations we try to get an answer depending on the position we choose.

# Measurable Benefits

<Describe the benefits in more detail>

Correct forecasts greatly help sales and help the company increase revenue. When we finished this model ,it can help the recommendation system function more usefully: Knowing the predicted price is very important for house recommendations for price-sensitive groups.

1. Project Detail

# Problem Description

<Describe the problem that is being addressed >

Find relation between price and house attributes.

# Identified Root Causes

<Describe the root cause(s) of the identified problem>

Humans can infer and predict the future from the historical information of the house. We want to find the relationship between the price and various attributes of the house from the data.

# Proposed Solution

<Describe the proposed solution using visuals, charts, etc>

We use machine learning to find the relationship of attributes of houses and price. It is important to understand the data, and it is important to clean and transform the data. There are both discrete and continuous features, and there are a lot of missing values. Fortunately, the contestant provided the file data\_description.txt, which describes the meaning of each feature. After understanding the content, most of the missing values can be smoothly interpolated.

If I have more time and more data, I want to try the neural network method. In the training process, as long as there are enough input x and output y, a better neural network model can be trained. This model is In similar housing price prediction problems, more accurate results can be obtained.

# Scope Details

|  |  |
| --- | --- |
| In Scope | Out of Scope |
| Download Kaggle House DataSet | Get more data rather than KaggleHouseData**Set** |
| Perform data preprocessing and cleaning | Analyze the meaning of each column |
| Apply machine learning to various models and algorithms | Create your own machine learning algorithms |
|  |  |
|  |  |
|  |  |

# Project Milestones

|  |  |  |
| --- | --- | --- |
| Name | Target Date | Comments |
| Project Kick off | 4.2 | Yes，we started at 4.1 |
| Milestone 2 | 4.5 | We finished data processing at that time |
| Milestone 3 | 4.9 | We finished main project at that time |
| Completion | 4.10 | We wrote all papers at 4.10 |

1. Project Costs & Resource Estimates

(INCLUDE HARDWARE AND SOFTWARE TO BE UTILIZED)

# Project Costs

|  |  |  |  |
| --- | --- | --- | --- |
| Project Expense | Cost Center | Comments | Est. Amount |
| A laptop | 1000 $ |  | 1000 $ |
| Data processing | 3-5 days |  | 3-5 working days |
| Turing the machine learning algorithm | 3-5days |  | 3-5 working days |
| Write the final report | 2-3 days |  | 2 working days |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Total:

# Resource Estimation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Role or Name | Department | Est. Hours | Rate | Est. Total |
| A laptop / a pc |  | 240 hours |  |  |
| 1 Developer (myself) |  | 240 hours |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Total:

1. Risks & Communication Plan

# Risk Mitigation Plan

|  |  |  |  |
| --- | --- | --- | --- |
| Identified Risk | Severity | Probability | Mitigation |
| Risk 1 | High | Low | Mitigation plans |
| Data process will take too long |  | Y | We choose open-source dataset, that relief the work of process data |
|  |  |  |  |
| Machine learning algorithm turing | Y |  | Use the combination method to find the optimal model combination and reduce programming time consumption |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Communication Plan

|  |  |  |  |
| --- | --- | --- | --- |
| Topic | Audience | Frequency/Date | Owner |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. Additional Project Details

<Provide any additional relevant project details here or links to supporting material>

Python is increasingly being used as a scientific language. Matrix and vector manipulations are extremely important for scientific computations. Both NumPy and Pandas have emerged to be essential libraries for any scientific computation, including machine learning, in python due to their intuitive syntax and high-performance matrix computation capabilities.

In doing the Kaggle predict match, We created a set of categorical features called categoricals that were not all included in the final model.

Go back and try to include these features. There are other methods that might help with categorical data, notably the ‘pandas’ method. After working on these features, repeat the transformations for the test data and make another try.

Working on models and participating in Kaggle competitions can be an iterative process — it’s important to experiment with new ideas, learn about the data, and test newer models and techniques.

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques>

<https://en.wikipedia.org/wiki/Principal_component_analysis>

<https://ruder.io/optimizing-gradient-descent/>