

DEVELOPMENT OF REAL ESTATE HOUSE PREDICTION DEMO APP FROM THE PROOF OF CONCEPT (POC) TO DEPLOYMENT STAGE



BY – JAY GUPTA (www.linkedin.com/in/jayguptanetwork)

CAR TECHNIQUE FOR SOLVING PROBLEMS

★ C A R ★

CAR Planning Technique for Solving Problems



CHALLENGE

State the problem or challenge clearly

- Define the problem as clearly as possible
- Present the written problem to Stakeholders to check if there is good alignment on the problem to be solved
- Follow the problem definition standard format capturing the context of the situation, problem statement, resources provided, timeline provided, resources provided, and metrics to evaluate the project success
- Focus on the expectations management with the Stakeholders as communication is one of the main keys to the project success



ACTION ITEMS

- Define an overview of the problem solving execution steps with timelines and state responsibilities for the team
- Focus on each step to clarify further how those steps will be done and who will be doing it if it is a team environment
- Focus on putting the execution steps in the chronological order so that it's easy to follow
- Focus on the list of tools to be used for the project so that necessary steps could be initiated by the right people at the right time
- Identify critical resources ahead of time so that a good risk mitigation plan could be put in place
- Document every step so that it's easy to follow and easy to explain to others as required



RESULTS EXPECTED

- Communication is the main key during the execution process and avoid any surprises
- Discuss the project progress regularly with the Stakeholders and keep sharing the project successes and challenges
- Ask for help, as needed on-time to avoid impacting the project timelines
- Review the budget clearly with the Stakeholders and don't assume anything
- Review the time required clearly with the Stakeholders
- Review the Success Metrics with the Stakeholders to avoid any confusion

PROBLEM HYPOTHESIS

Problem Statement Worksheet (Hypothesis Formation)

I
Development of Real Estate House Prediction App from the Proof of Concept (PoC) to Product Deployment Stage for Private Demonstration by end of June, 2021

H

1 Context

The problem is based on start-up companies where there are no data engineers or any other support team to deploy any machine learning application. Also start-up environment demands that app is developed quickly with no cost up front using open sources and freely available platforms to establish the Proof of Concept before spending any money.

2 Criteria for success

1. Use only freely available resources including the publicly available dataset
2. Develop a working application for private demonstration purpose only so focus should be on the speed of deployment and keep it simple to deploy.
3. Basic interface development only for demo purpose

3 Scope of solution space

Using the publicly available dataset, focus should be on developing the real estate house pricing prediction application which can be deployed quickly without too much extra coding.

4 Constraints within solution space

1. No paid platforms are to be used.
2. No paid dataset is to be utilized as the basic focus on developing the concept and this is not the final application.
3. Only freely available platforms are to be used for hosting the application.

5 Stakeholders to provide key insight

1. Discuss with mentor on a regular basis
2. Present the final working model to mentor

6 Key data sources

Publicly available house prediction dataset at Kaggle is the most advanced dataset available and it is proposed to be used for this application.

EXECUTION METHODOLOGY

- Follow the Data Science Methodology (DSM) for the machine learning model development
- Create an architecture model to keep the management team in the loop with the project developments
- Create an app deployment framework to show the deployment path clearly to various stakeholders
- Follow the "STAR" methodology to capture the documentation and share the insights for future use

DSM METHODOLOGY

01 PROBLEM IDENTIFICATION 02 DATA WRANGLING 03 EXPLORATORY DATA ANALYSIS 04 PRE-PROCESSING AND TRAINING DATA DEVELOPMENT 05 MODELING 06 DOCUMENTATION



Identify the correct problem to solve



Collect, organize, define, and clean a relevant dataset



Understand the relationship between data and features



Standardize and train your dataset



Select, train and deploy a model to make predictive insights



Document your work and share your findings

ARCHITECTURE DIAGRAM

★ ARCHITECTURE DIAGRAM ★

For House Price App Project



Jupyter Notebook

- Use Jupyter Notebooks for step-by-step analysis from problem identification to deployment stage
- Capture all steps for the documentation purpose
- Capture presentation for future use in the final stage of the notebook



PyCharm IDE

- Since the production environment coding is different than the Jupyter Notebook coding, use PyCharm IDE for the coding development.
- Keep the coding ready in case the application is required for the user testing pilot project.
- Create the model pickle file for use during the next step of application development



Streamlit App

- Use command line interface (CLI) to create the virtual environment for the application deployment
- Create files such as app.py, EDA_app.py and ML_py for the application purpose in a text editor like Sublime
- Keep testing the application via Streamlit interface to check its working and connectivity with each app element



GitHub Repository

- Create repository on the GitHub to containerize the application with all the required files
- Clone the repository via CLI and push all the required files to GitHub

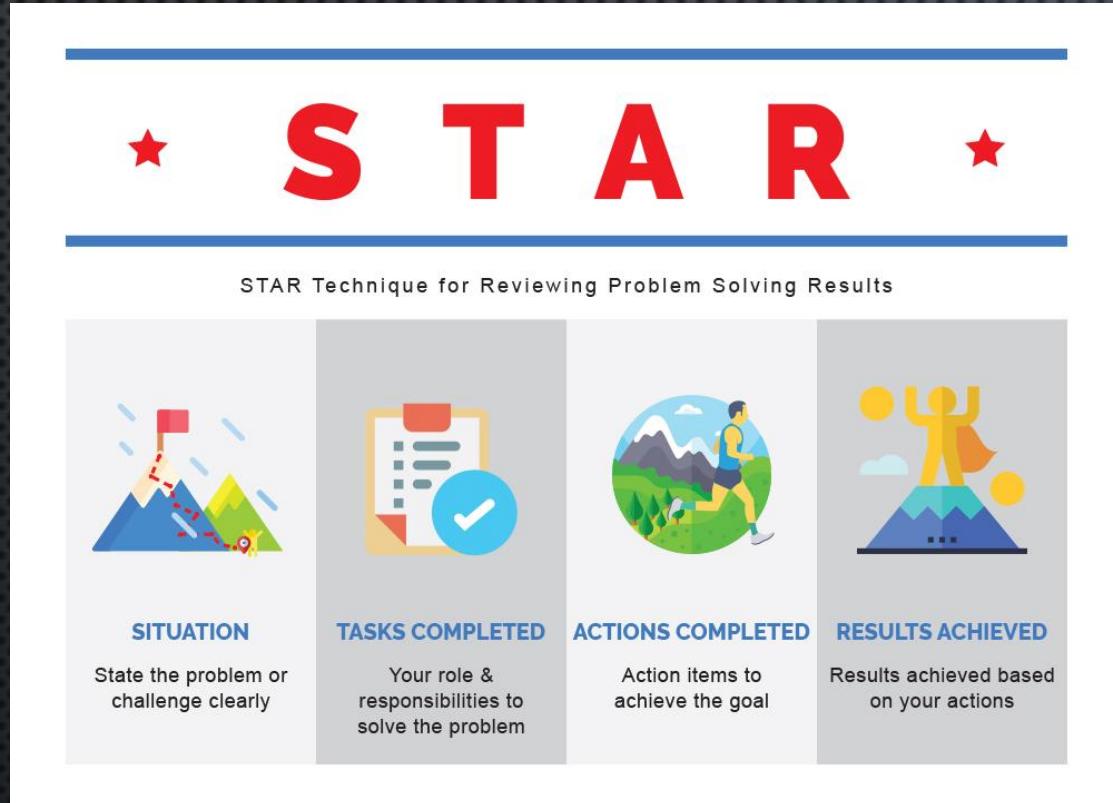


Heroku Platform

- Connect to Heroku via CLI after signing up at the Heroku platform
- Connect GitHub repository to Heroku platform
- Create a name for your app and then take the URL as provided by Heroku
- Push all the contents of GitHub to Heroku for the app deployment

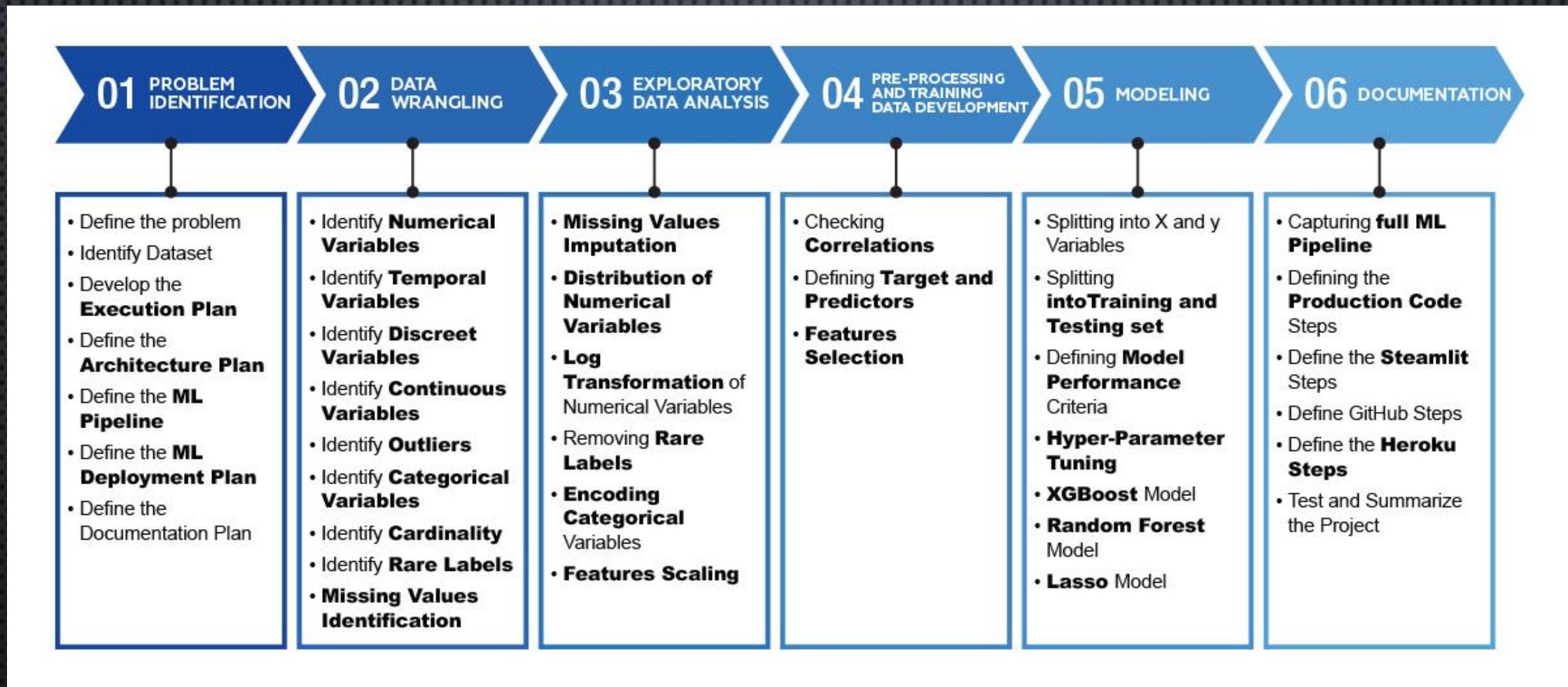
(Image Credits: <https://jupyter.org>, <https://www.jetbrains.com/pycharm>, <https://blog.streamlit.io>, <https://github.com>, <https://www.heroku.com>)

STAR METHODOLOGY FOR KEY INSIGHTS

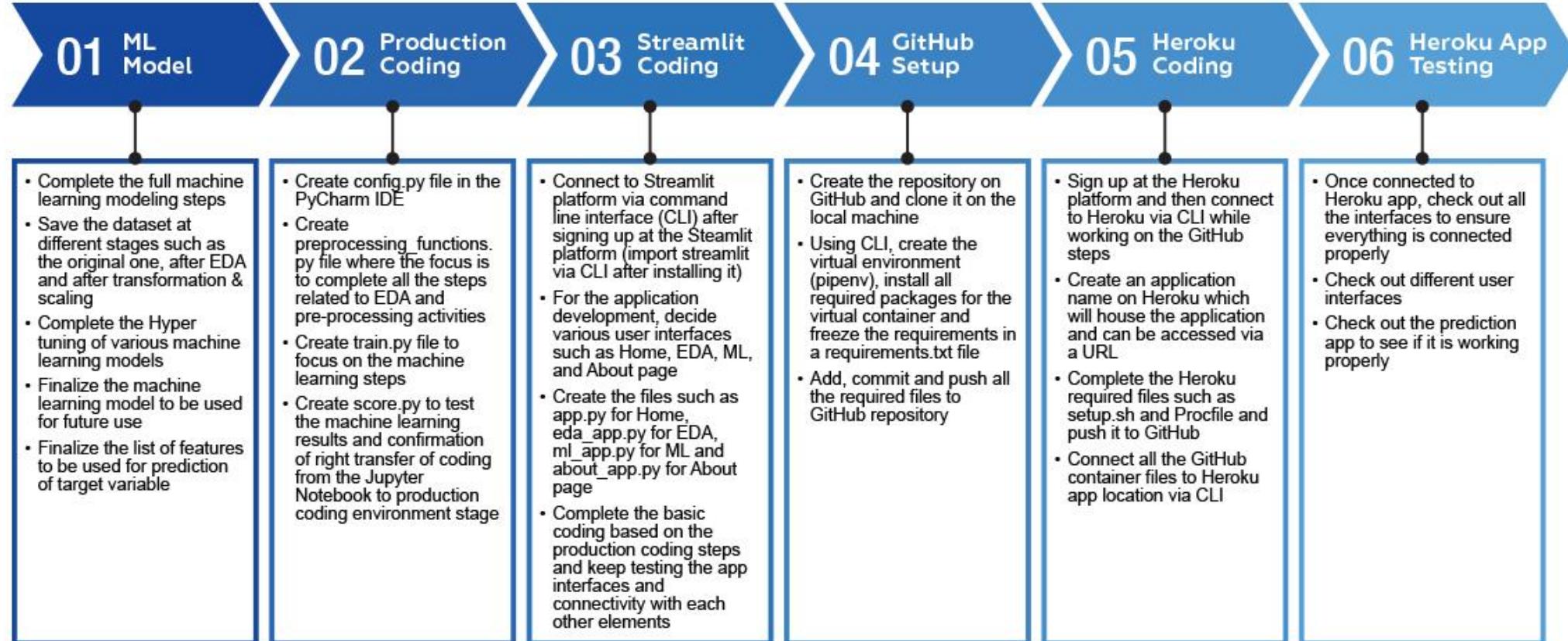


- Use publicly available free datasets to experiment and focus on the feasibility projects
- It is possible to do experiments for the real-world problems without a big or no budget
- Focus on the speed rather than perfecting the model as these projects are designed to prove the proof of concept
- Use Streamlit like apps to make the machine learning model explainable and fit for presentation to various stakeholders
- Follow the Minimum Viability Product criteria while doing these type of projects

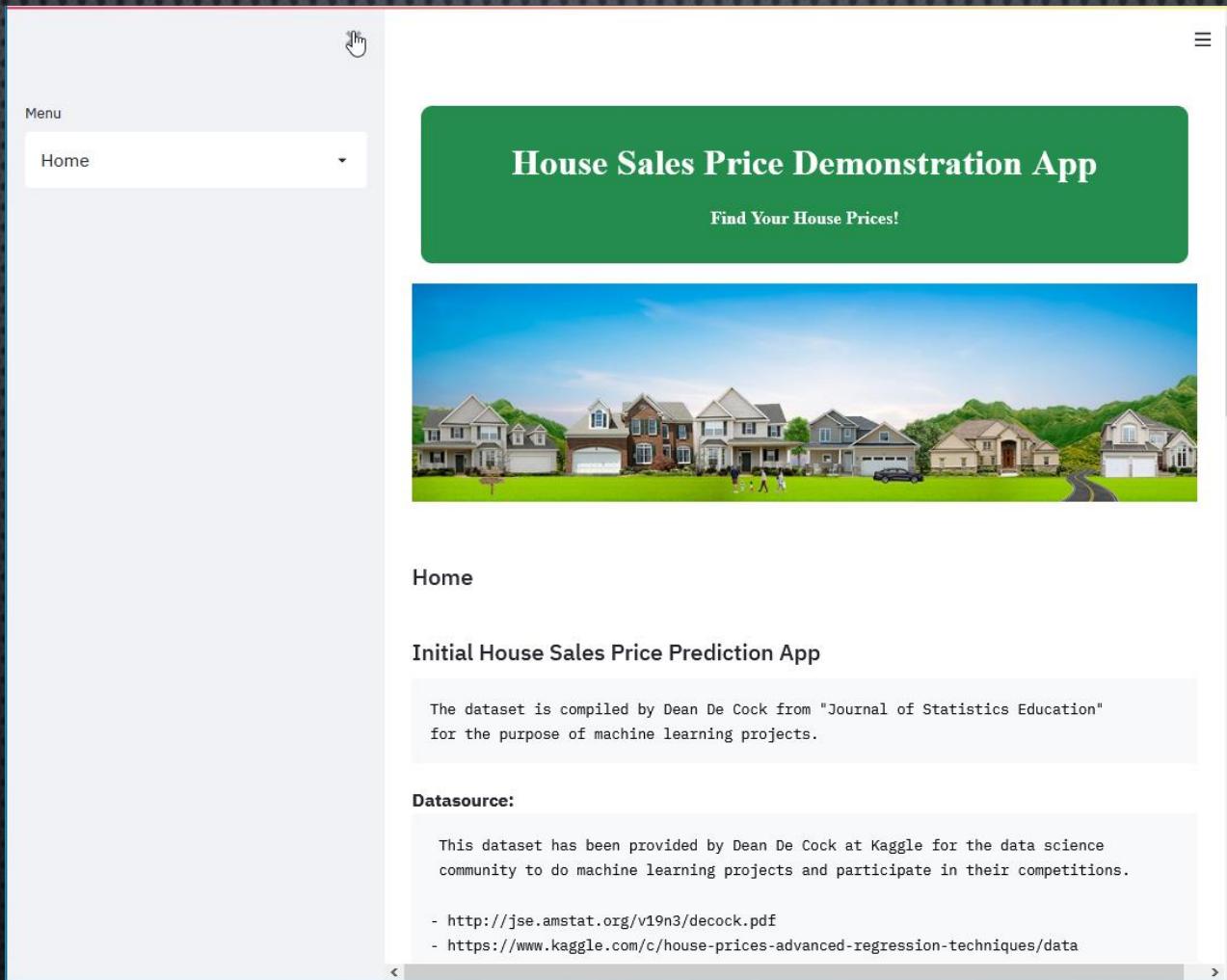
MODELING ACTIVITIES SUMMARY



DEPLOYMENT ACTIVITIES SUMMARY



APP ON HEROKU – HOME PAGE



Access App on Heroku at <https://housepricedemoapp.herokuapp.com/>

APP ON HEROKU – EDA PAGE # 1

House Sales Price Demonstration App

Find Your House Prices!



Running EDA Application

Original Data before EDA +

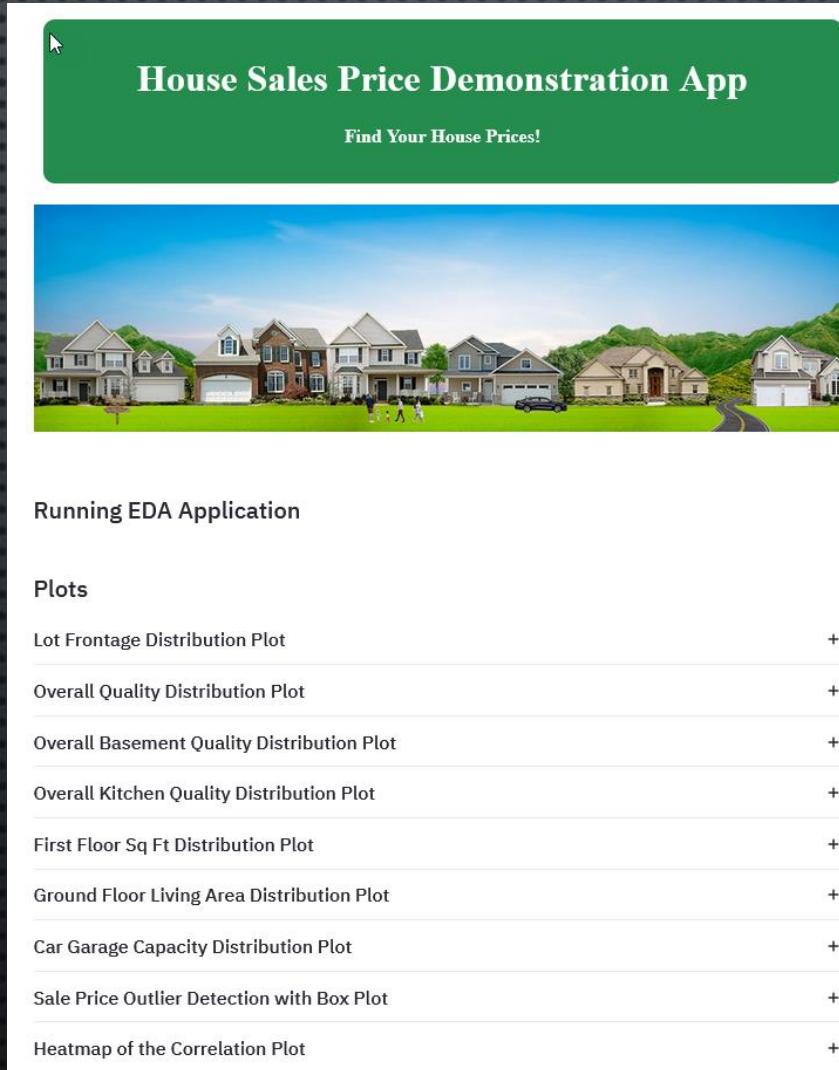
Data after Cleaning, Imputing, Feature Selection and Encoding +

Data Types after Pre-Processing Steps +

Descriptive Data Summary before ML +

Access App on Heroku at <https://housepricedemoapp.herokuapp.com/>

APP ON HEROKU – EDA PAGE # 2



Access App on Heroku at <https://housepricedemoapp.herokuapp.com/>

APP ON HEROKU – ML PAGE



House Sales Price Demonstration App

Find Your House Prices!



Machine Learning Modeling

Click on the Plus Sign to Select the Options on Community Features: +

Click on the Plus Sign to Provide Values for the House: +

Click on the Plus Sign to Select House Feature Options: +

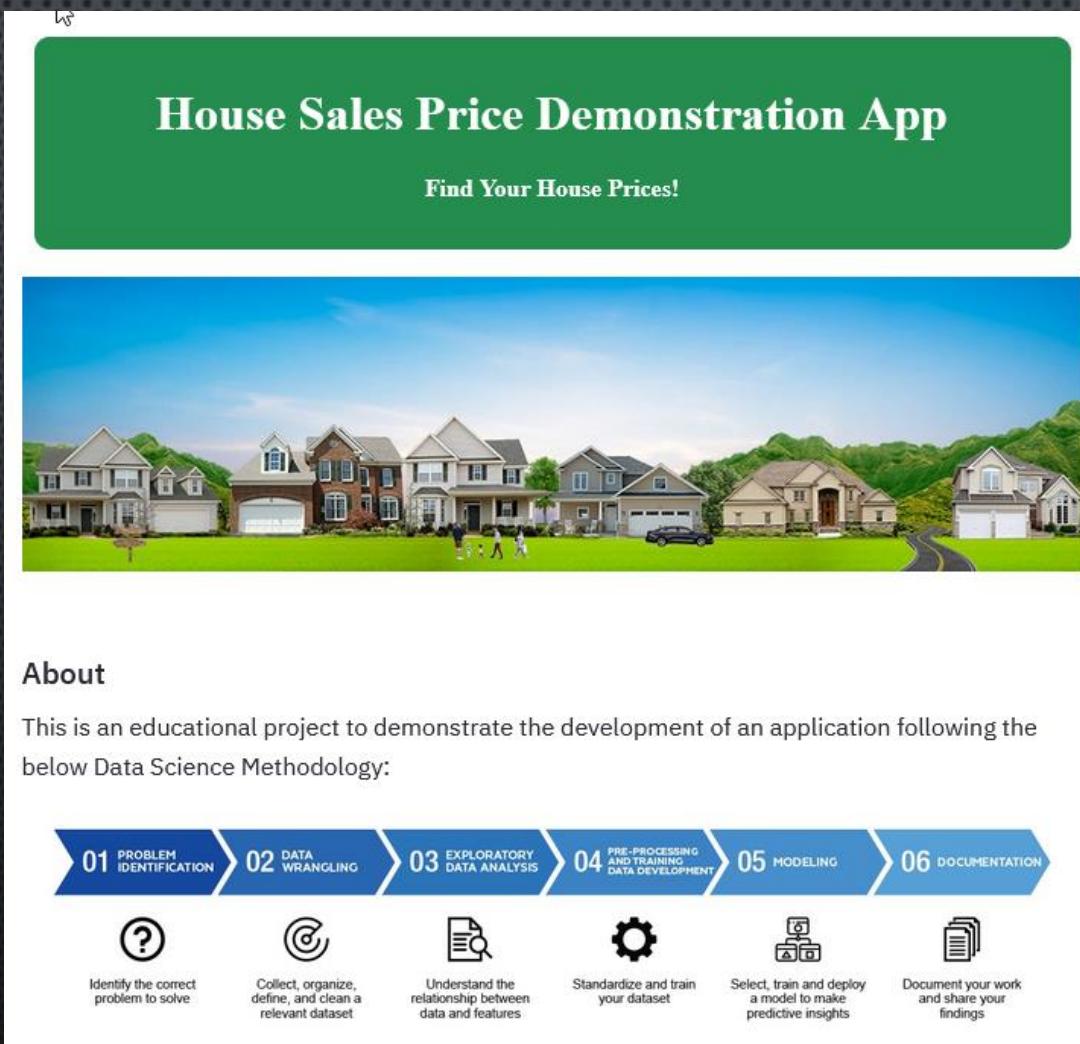
Click on the Plus Sign to Select House Exterior Options: +

Your Selected Options +

Predicted House Prices +

Access App on Heroku at <https://housepricedemoapp.herokuapp.com/>

APP ON HEROKU – ABOUT PAGE



Access App on Heroku at <https://housepricedemoapp.herokuapp.com/>

THANK YOU

BY – JAY GUPTA

LINKEDIN: [HTTPS://WWW.LINKEDIN.COM/IN/JAYGUPTANETWORK/](https://www.linkedin.com/in/jayguptanetwork/)

GITHUB: [HTTPS://JAYGUPTACAL.GITHUB.IO/PORTFOLIO/](https://jayguptacal.github.io/portfolio/)