# Data Analysis Project

## Data Analyst: Manoj Menon

## Client/Sponsor: Lead Data Scientist (Manager)

Other Stakeholders : Analytics team

## Purpose:

Goal is to develop a best fitting model that can predict the car price based on relevant features.

*Inspect the dataset that is available, to analyze key features (that were already identified earlier) and use it to develop a model for prediction of car prices. Develop linear and non-linear regression models. Evaluate the accuracy and appropriateness of each model type to find the best fit model.*

## Scope / Major Project Activities:

*What are the major parts of this project? List out the high-level steps, activities, or stages of the project, and give a brief description for each.*

|  |  |
| --- | --- |
| Activity | Description |
| Collect required data | Examine available dataset |
| Build a SLR model | Construct a Simple Linear Regression model for a single predictor variable |
| Build a MLR model | Construct a Multiple Linear regression model using multiple predictor variables |
| Evaluate the models visually | Use Regression plots and Residual plot to evaluate the SLR  Use Distribution plots to evaluate the MLR |
| Build a Polynomial model | Use Pipelines to create and test a non-linear regression model |
| Evaluate the models numerically | Use R^2 and MSE to evaluate and compare the various model results. |
| Recommend strategy | Recommend which model is the best fit for the prediction |

## This project does not include:

* This project is primarily targeted towards learning to apply Regression models. Other model strategies are not considered here

## Deliverables:

*A specific list of things that your project will deliver.*

|  |  |
| --- | --- |
| Deliverable | Description/ Details |
| Jupyter Notebook | The complete project will be documented using a Jupyter Notebook |
| Recommendations for modelling | Conclusions from the numerical evaluations of the various model approaches and the best model to use |
|  |  |

## Schedule Overview / Major Milestones:

*The expected schedule for the project. This can be defined by milestones (e.g. “all data is cleaned and processed”), periods of time (“Week 1 / Week 2”), or other ways based on the needs of the project.*

|  |  |  |
| --- | --- | --- |
| Milestone | Expected Completion Date | Description/Details |
| *Data collection* | *May 5th 2022* | *Acquire required data* |
| *Prepare Data for Review* | *May 6th 2022* | *Organize and inspect data* |
| *Model creation* | *May 8th 2022* | *Create all data models* |
| *Model evaluation* | *May 10th 2022* | *Visual and numeric evaluation with results* |
| *Recommendations* | *May 11th 2022* | *Main model recommendations are prepared* |

## \*Estimated date for completion:

*May 20th including revisions after feedback.*