1) Arina Karpova, Sara Morcillo, Nicola Capeto

2) The goal is to build a machine learning model to predict the selling prices of houses based on a variety of features on which the value of the house is evaluated. It is a modelling project

3) Brief description of the dataset

**Data**: The data set consists of information on some 22,000 properties.  The dataset consisted of historic data of houses sold between May 2014 to May 2015. These are the definitions of data points provided.

* **Id**: Unique identification number for the property.
* **date**: date the house was sold.
* **price**: price of the house.
* **waterfront**: house which has a view to a waterfront.
* **condition**: How good the condition is (overall). **1** indicates worn out property and **5** excellent.
* **grade**: Overall grade given to the housing unit, based on King County grading system. 1 poor ,13 excellent.
* **Sqft\_above**: square footage of house apart from basement.
* **Sqft\_living15**: Living room area in 2015(implies - some renovations). This might or might not have affected the lotsize area.
* **Sqft\_lot15**: lotSize area in 2015(implies - some renovations).

4) High-level Planning of the project:

Day 1 – EDA, Cleaning

Day 2 - Cleaning the data, remove typos, correcting datatypes, plotting correlation map, eliminating correlated columns

Day 3- Transform the data, Scale Numerical Feature, Encode Categorical Features

Day 4 – Modelling, Fitting, Evaluation, Validation, Improving and Visualization