PROJECT DEVELOPMENT PHASE
(DELIVERY OF SPRINT-1)
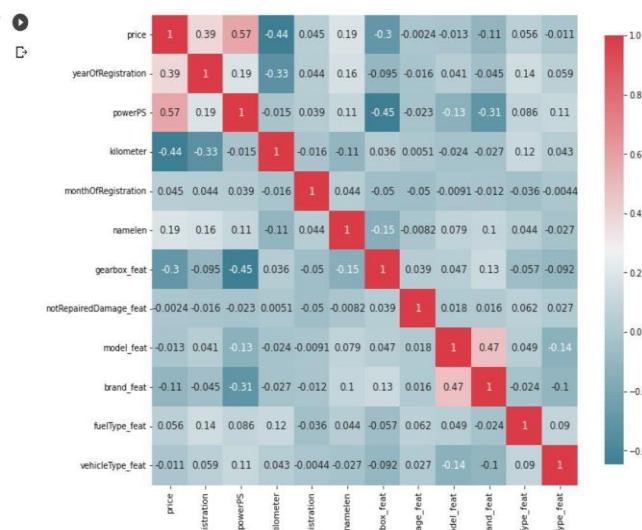
Date	**10th November 2022**
Team ID	**PNT2022TMID01716**
Project Name	**Car Resale Value prediction**
Maximum marks	**4 Marks**

Import Library and load the data set:

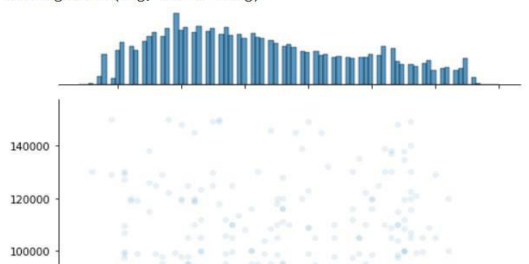## Understanding and analyzing the dataset by Correlation:

Clearing the null values:

```
[ ]  dedups['notRepairedDamage'].fillna(value='not-declared', inplace=True)
     dedups['fuelType'].fillna(value='not-declared', inplace=True)
     dedups['gearbox'].fillna(value='not-declared', inplace=True)
     dedups['vehicleType'].fillna(value='not-declared', inplace=True)
     dedups['model'].fillna(value='not-declared', inplace=True)
```

```
dedups['namelen'] = [min(70, len(n)) for n in dedups['name']]

ax = sns.jointplot(x='namelen',
                   y='price',
                   data=dedups[['namelen','price']],
#                  data=dedups[['namelen','price']][dedups['model']=='golf'],
                   alpha=0.1,
                   size=8)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:2182: UserWarning: The `size` parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
```


Preprocessing the Categorical values:

```
labels = ['name', 'gearbox', 'notRepairedDamage', 'model', 'brand', 'fuelType', 'vehicleType']
les = {}

for l  Loading...
    les[l] = preprocessing.LabelEncoder()
    les[l].fit(dedups[l])
    tr = les[l].transform(dedups[l])
    dedups.loc[:, l + '_feat'] = pd.Series(tr, index=dedups.index)

labeled = dedups[ ['price'
                  ,'yearOfRegistration'
                  ,'powerPS'
                  ,'kilometer'
                  ,'monthOfRegistration'
                  , 'namelen']
                 + [x+"_feat" for x in labels]]
```