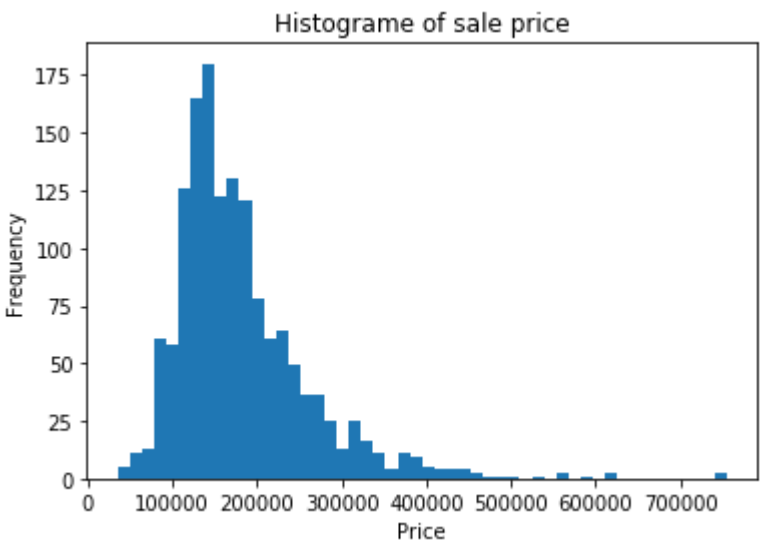
[1]
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os
import sys
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
%matplotlib inline
```

[2]
```
data = pd.read_csv('../train.csv', index_col=0)
data.head()
```

	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	
Id							
1	60	RL	65.0	8450	Pave	NaN	
2	20	RL	80.0	9600	Pave	NaN	
3	60	RL	68.0	11250	Pave	NaN	
4	70	RL	60.0	9550	Pave	NaN	
5	60	RL	84.0	14260	Pave	NaN	

5 rows × 80 columns

[5]
```
plt.hist(data.SalePrice, bins=50)
plt.xlabel('Price')
plt.ylabel('Frequency')
plt.title('Histograme of sale price');
```



[3]
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1460 entries, 1 to 1460
Data columns (total 80 columns):
MSSubClass       1460 non-null int64
MSZoning         1460 non-null object
LotFrontage      1201 non-null float64
LotArea          1460 non-null int64
Street           1460 non-null object
Alley            91 non-null object
LotShape         1460 non-null object
LandContour      1460 non-null object
Utilities        1460 non-null object
LotConfig        1460 non-null object
LandSlope        1460 non-null object
Neighborhood     1460 non-null object
Condition1       1460 non-null object
Condition2       1460 non-null object
BldgType         1460 non-null object
HouseStyle       1460 non-null object
OverallQual      1460 non-null int64
OverallCond      1460 non-null int64
YearBuilt        1460 non-null int64
YearRemodAdd     1460 non-null int64
RoofStyle        1460 non-null object
RoofMatl         1460 non-null object
Exterior1st      1460 non-null object
Exterior2nd      1460 non-null object
MasVnrType       1452 non-null object
MasVnrArea       1452 non-null float64
ExterQual        1460 non-null object
ExterCond        1460 non-null object
Foundation       1460 non-null object
BsmtQual         1423 non-null object
```

[4]

```
data.Alley = data.Alley.fillna(value = 'NoAlley')
data.BsmtCond = data.BsmtCond.fillna(value = 'NoBsmt')
data.BsmtQual = data.BsmtQual.fillna(value = 'NoBsmt')
data.BsmtExposure = data.BsmtExposure.fillna(value= 'NoBsmt')
data.BsmtFinType1 = data.BsmtFinType1.fillna(value= 'NoBsmt')
data.BsmtFinType2 = data.BsmtFinType2.fillna(value= 'NoBsmt')
data.LotFrontage = data.LotFrontage.fillna(value = 0)
data.FireplaceQu = data.FireplaceQu.fillna(value = 'Nofireplace')
data.GarageType = data.GarageType.fillna(value = 'NoGarage')
data.GarageCond = data.GarageCond.fillna(value = 'NoGarage')
data.GarageFinish = data.GarageFinish.fillna(value = 'NoGarage')
data.GarageYrBlt = data.GarageYrBlt.fillna(value = 0)
data.GarageQual = data.GarageQual.fillna(value = 'NoGarage')

data.PoolQC = data.PoolQC.fillna(value = 'NoPool')
data.Fence = data.Fence.fillna(value = 'NoFence')
data.MiscFeature = data.MiscFeature.fillna(value = 'NoMisc')
data.MasVnrType = data.MasVnrType.fillna(value = 'noMas')
data.MasVnrArea = data.MasVnrArea.fillna(value = 'noMas')

data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1460 entries, 1 to 1460
Data columns (total 80 columns):
MSSubClass       1460 non-null int64
MSZoning         1460 non-null object
LotFrontage      1460 non-null float64
LotArea          1460 non-null int64
Street           1460 non-null object
Alley            1460 non-null object
LotShape         1460 non-null object
LandContour      1460 non-null object
Utilities        1460 non-null object
LotConfig        1460 non-null object
LandSlope        1460 non-null object
Neighborhood     1460 non-null object
Condition1       1460 non-null object
Condition2       1460 non-null object
BldgType         1460 non-null object
HouseStyle       1460 non-null object
OverallQual      1460 non-null int64
OverallCond      1460 non-null int64
YearBuilt        1460 non-null int64
YearRemodAdd     1460 non-null int64
RoofStyle        1460 non-null object
RoofMatl         1460 non-null object
Exterior1st      1460 non-null object
Exterior2nd      1460 non-null object
MasVnrType       1460 non-null object
MasVnrArea       1460 non-null object
ExterQual        1460 non-null object
ExterCond        1460 non-null object
Foundation       1460 non-null object
BsmtQual         1460 non-null object
```

[9]
```
n_data = pd.get_dummies(data)
```

[10]
```
n_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1460 entries, 1 to 1460
Columns: 631 entries, MSSubClass to SaleCondition_Partial
dtypes: float64(2), int64(34), uint8(595)
memory usage: 1.2 MB
```

[11]
```
n_data.head()
```

	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	Ye
Id						
1	60	65.0	8450	7	5	20
2	20	80.0	9600	6	8	19

	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	Ye
Id						
3	60	68.0	11250	7	5	20
4	70	60.0	9550	7	5	19
5	60	84.0	14260	8	5	20

5 rows × 631 columns

```
[ ]  n_data.to_csv('../clean_data.csv')
```

```
[ ]  data.to_csv('../semi_clean_data.csv')
```