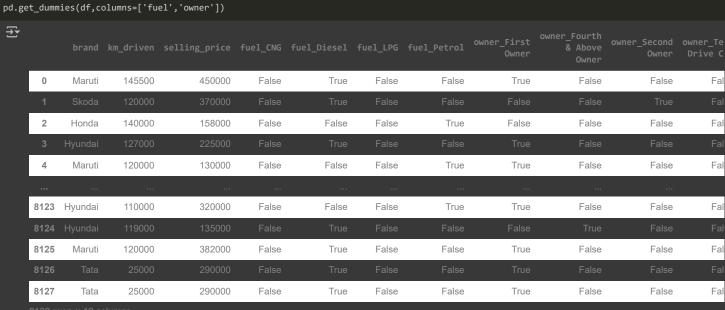


1. OneHotEncoding using Pandas



2. K-1 OneHotEncoding

.get_dummies(df,columns=['fuel','owner'],drop_first=True)											
	brand		selling_price	fuel_Diesel	fuel_LPG	fuel_Petrol	owner_Fourth & Above Owner	owner_Second Owner	owner_Test Drive Car	owner_Third Owner	
) Maruti	145500	450000	True	False	False	False	False	False	False	
		120000	370000								
2	2 Honda	140000	158000	False	False	True	False	False	False	True	
;	B Hyundai	127000	225000	True	False	False	False	False	False	False	
4	l Maruti	120000	130000	False	False	True	False	False	False	False	
81	23 Hyundai	110000	320000	False	False	True	False	False	False	False	
81	24 Hyundai	119000	135000	True	False	False	True	False	False	False	
81	25 Maruti	120000	382000	True	False	False	False	False	False	False	
81	26 Tata	25000	290000	True	False	False	False	False	False	False	
81	27 Tata	25000	290000	True	False	False	False	False	False	False	
4											>

9/11/24, 12:57 PM ONE_-HOT_ENCODING.ipynb - Colab Becaue during ML project we does not use pandas because it does not remember the coloumn name which we skip or use. So during ml project 3. OneHotEncoding using Sklearn from sklearn.model_selection import train_test_split X_train,X_test,y_train,y_test = train_test_split(df.iloc[:,0:4],df.iloc[:,-1],test_size=0.2,random_state=2) X_train.head() **₹** 5571 Hyundai 35000 Diesel First Owner 2957 Hyundai 25000 Petrol First Owner First Owner 6684 Hyundai 155000 Diesel Next steps: Generate code with X_train View recommended plots New interactive sheet from sklearn.preprocessing import OneHotEncoder ohe = OneHotEncoder(drop='first',sparse=False,dtype=np.int32) # ohe = OneHotEncoder() X_train_new = ohe.fit_transform(X_train[['fuel','owner']]) X_train_new 🚁 /usr/local/lib/python3.10/dist-packages/sklearn/preprocessing/_encoders.py:975: FutureWarning: `sparse` was renamed to `sparse_output` warnings.warn(4 X_test_new = ohe.transform(X_test[['fuel','owner']]) X_train_new.shape \rightarrow (6502, 7) np.hstack((X_train[['brand','km_driven']].values,X_train_new)) → array([['Hyundai', 35000, 1, ..., 0, 0, 0], 'Jeep', 60000, 1, ..., 0, 0, 0], ['Hyundai', 25000, 0, ..., 0, 0, 0],