```
import pandas as pd
import numpy as np


from scipy import stats

def ttest_categorical_numerical(categorical_col, numerical_col, data):

    # Split the data into two groups based on the categorical variable
    group1 = data[numerical_col][data[categorical_col] == data[categorical_col].unique()[0]]
    group2 = data[numerical_col][data[categorical_col] == data[categorical_col].unique()[1]]

    # Perform t-test
    t_statistic, p_value = stats.ttest_ind(group1, group2, equal_var=False)

    if(p_value < 0.05):
      return 1,p_value
    else:
      return 0,p_value



from scipy.stats import chi2_contingency

def chi2_categorical_categorical(categorical_col1, categorical_col2, data):

    # Create a contingency table
    contingency_table = pd.crosstab(data[categorical_col1], data[categorical_col2])

    # Perform chi-squared test
    chi2_statistic, p_value, degrees_of_freedom, expected_values = chi2_contingency(contingency_table)

    if(p_value < 0.05):
      return 1,p_value
    else:
      return 0,p_value


from scipy.stats import f_oneway

def anova_categorical_numerical(categorical_col, numerical_col, data):

    # Get unique categories in categorical column
    categories = data[categorical_col].unique()

    # Create empty lists for each group's numerical data
    group_data = [data[numerical_col][data[categorical_col] == category] for category in categories]

    # Perform one-way ANOVA
    f_statistic, p_value = f_oneway(*group_data)

    # Check significance level
    if p_value < 0.05:
        return 1,p_value
    else:
        return 0,p_value


df = pd.read_csv("/content/drive/MyDrive/Datasets/hr data.csv")
data = df.copy()
data.columns
```

```
Index(['SLNO', 'Candidate Ref', 'DOJ Extended', 'Duration to accept offer',
       'Notice period', 'Offered band', 'Pecent hike expected in CTC',
       'Percent hike offered in CTC', 'Percent difference CTC',
       'Joining Bonus', 'Candidate relocate actual', 'Gender',
       'Candidate Source', 'Rex in Yrs', 'LOB', 'Location', 'Age', 'Status'],
      dtype='object')
```

```
data.drop(["SLNO","Candidate Ref"],axis = 1,inplace = True)


independent_cols = data.columns[:-1]
hypothesis_flag = [0]*len(independent_cols)
relevant_col = []
target_col = data.columns[-1]


for index,col in enumerate(independent_cols):
  print("Null Hypothesis: There is no significant relationship between "+col+" and candidate "+target_col)
  print("Alternative Hypothesis: There is significant relationship between "+col+" and candidate"+target_col)
```

```
    if np.issubdtype(data[col].dtype, np.number):
      flag,p_value = ttest_categorical_numerical(target_col,col,data)
      if(flag == 1):
        print("Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance")
        print(p_value)
        print("\n")
        hypothesis_flag[index] = 1
        relevant_col.append(col)
      else:
        print("Failed to reject null hypothesis, no significant relationship exists at 0.05 level of significance")
        print(p_value)
        print("\n")

    else:
      flag,p_value = chi2_categorical_categorical(col,target_col,data)
      if(flag == 1):
        print("Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance")
        print(p_value)
        print("\n")
        hypothesis_flag[index] = 1
        relevant_col.append(col)
      else:
        print("Failed to reject null hypothesis, no significant relationship exists at 0.05 level of significance")
        print(p_value)
        print("\n")
```

```
Null Hypothesis: There is no significant relationship between DOJ Extended and candidate Status
Alternative Hypothesis: There is significant relationship between DOJ Extended and candidateStatus
Failed to reject null hypothesis, no significant relationship exists at 0.05 level of significance
0.5811864020628462


Null Hypothesis: There is no significant relationship between Duration to accept offer and candidate Status
Alternative Hypothesis: There is significant relationship between Duration to accept offer and candidateStatus
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance
8.116770482582802e-08


Null Hypothesis: There is no significant relationship between Notice period and candidate Status
Alternative Hypothesis: There is significant relationship between Notice period and candidateStatus
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance
3.5646231564497353e-64


Null Hypothesis: There is no significant relationship between Offered band and candidate Status
Alternative Hypothesis: There is significant relationship between Offered band and candidateStatus
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance
0.03537171423302974


Null Hypothesis: There is no significant relationship between Pecent hike expected in CTC and candidate Status
Alternative Hypothesis: There is significant relationship between Pecent hike expected in CTC and candidateStatus
Failed to reject null hypothesis, no significant relationship exists at 0.05 level of significance
0.9761873982800147


Null Hypothesis: There is no significant relationship between Percent hike offered in CTC and candidate Status
Alternative Hypothesis: There is significant relationship between Percent hike offered in CTC and candidateStatus
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance
0.0042693259375661775


Null Hypothesis: There is no significant relationship between Percent difference CTC and candidate Status
Alternative Hypothesis: There is significant relationship between Percent difference CTC and candidateStatus
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance
0.0017796484128319533


Null Hypothesis: There is no significant relationship between Joining Bonus and candidate Status
Alternative Hypothesis: There is significant relationship between Joining Bonus and candidateStatus
Failed to reject null hypothesis, no significant relationship exists at 0.05 level of significance
0.7454707857023846


Null Hypothesis: There is no significant relationship between Candidate relocate actual and candidate Status
Alternative Hypothesis: There is significant relationship between Candidate relocate actual and candidateStatus
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance
5.352550694431075e-77


Null Hypothesis: There is no significant relationship between Gender and candidate Status
Alternative Hypothesis: There is significant relationship between Gender and candidateStatus
Failed to reject null hypothesis, no significant relationship exists at 0.05 level of significance
0.23679896899569233
```

```
numerical_col = [col for col in independent_cols if (np.issubdtype(data[col].dtype, np.number) and col in relevant_col)]
categorical_col = [col for col in independent_cols if (np.issubdtype(data[col].dtype, np.number) != True and col in relevant_col)]
```

```
numerical_col
```

```
['Duration to accept offer',
 'Notice period',
 'Percent hike offered in CTC',
 'Percent difference CTC',
 'Rex in Yrs',
 'Age']
```

```
categorical_col
```

```
['Offered band',
 'Candidate relocate actual',
 'Candidate Source',
 'LOB',
 'Location']
```

```
for cat_col in categorical_col:
  for num_col in numerical_col:

    if(data[cat_col].nunique() == 2):
      flag,p_value = ttest_categorical_numerical(cat_col,num_col,data)
      if(flag == 1):
        print("Null Hypothesis: There is no significant relationship between "+cat_col+" and  "+num_col)
        print("Alternative Hypothesis: There is significant relationship between "+cat_col+" and "+num_col)
        print(p_value)
        print("Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance")
        print("\n")
    else:
      flag,p_value = anova_categorical_numerical(cat_col,num_col,data)
      if(flag == 1):
        print("Null Hypothesis: There is no significant relationship between "+cat_col+" and  "+num_col)
        print("Alternative Hypothesis: There is significant relationship between "+cat_col+" and "+num_col)
        print(p_value)
        print("Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance")
        print("\n")
```

```
Null Hypothesis: There is no significant relationship between Offered band and  Duration to accept offer
Alternative Hypothesis: There is significant relationship between Offered band and Duration to accept offer
5.464186363937911e-16
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance


Null Hypothesis: There is no significant relationship between Offered band and  Notice period
Alternative Hypothesis: There is significant relationship between Offered band and Notice period
8.915504570908882e-55
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance


Null Hypothesis: There is no significant relationship between Offered band and  Percent hike offered in CTC
Alternative Hypothesis: There is significant relationship between Offered band and Percent hike offered in CTC
0.00847022361648555
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance


Null Hypothesis: There is no significant relationship between Offered band and  Percent difference CTC
Alternative Hypothesis: There is significant relationship between Offered band and Percent difference CTC
5.067879054206044e-43
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance


Null Hypothesis: There is no significant relationship between Offered band and  Rex in Yrs
Alternative Hypothesis: There is significant relationship between Offered band and Rex in Yrs
0.0
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance


Null Hypothesis: There is no significant relationship between Offered band and  Age
Alternative Hypothesis: There is significant relationship between Offered band and Age
0.0
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance


Null Hypothesis: There is no significant relationship between Candidate relocate actual and  Notice period
Alternative Hypothesis: There is significant relationship between Candidate relocate actual and Notice period
6.664321247113422e-10
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance


Null Hypothesis: There is no significant relationship between Candidate relocate actual and  Age
Alternative Hypothesis: There is significant relationship between Candidate relocate actual and Age
```

```
1.6884966477999466e-07
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance


Null Hypothesis: There is no significant relationship between Candidate Source and  Duration to accept offer
Alternative Hypothesis: There is significant relationship between Candidate Source and Duration to accept offer
2.0576076819652775e-05
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance


Null Hypothesis: There is no significant relationship between Candidate Source and  Notice period
Alternative Hypothesis: There is significant relationship between Candidate Source and Notice period
0.0029057345720703356
Alternate Hypothesis accepted, significant relationship exists at 0.05 level of significance
```

✓ 0s completed at 11:20 PM