Practical 4 Use the dataset named "People Charm case.csv" that deals with HR analytics.

Answer the following questions:

1. Which of the variables have missing values?
2. What is the third quartile value for the variable "lastEvaluvation"?
3. Construct a Crosstable for the variables 'dept' and "salary" and find out which department has highest frequency value in the category low salary.
4. Generate a boxplot for the variable "numberOfProjects" and get the median value for the number of projects where the employees have worked on.
5. Plot a histogram using the variable "avgMonthlyHours" and find the range in which the number of employees worked for 150 hours per month?
6. Generate a boxplot for the variables "lastEvaluation" and "numberOfProjects"

1. Which of the variables have missing values?

```
import pandas as pd
# Load the dataset
df = pd.read_csv("People Charm case.csv")
# Check for missing values
missing_values = df.isnull().sum()
variables_with_missing = missing_values[missing_values > 0].index.tolist()
print("Variables with missing values:")
print(variables_with_missing)
```

2. What is the third quartile value for the variable "lastEvaluvation"?

```
# Assuming 'lastEvaluvation' is a typo and should be 'lastEvaluation'
third_quartile = df['lastEvaluation'].quantile(0.75)
print("Third quartile value for lastEvaluation:", third_quartile)
```

3. Construct a Crosstable for the variables 'dept' and "salary" and find out which department has highest frequency value in the category low salary.

```
# Crosstab and finding the department with highest frequency of low salary
cross_table = pd.crosstab(df['dept'], df['salary'])

# Find department with highest frequency of low salary
dept_with_highest_low_salary = cross_table['low'].idxmax()
highest_freq_low_salary = cross_table.loc[dept_with_highest_low_salary, 'low']

print("Department with highest frequency of low salary:", dept_with_highest_low_salary)
print("Frequency of low salary:", highest_freq_low_salary)
```

4. Generate a boxplot for the variable "numberOfProjects" and get the median value for the number of projects where the employees have worked on.

```
import matplotlib.pyplot as plt

# Boxplot for numberOfProjects
plt.figure(figsize=(8, 6))
plt.boxplot(df['numberOfProjects'])
plt.title('Boxplot of numberOfProjects')
plt.ylabel('Number of Projects')
plt.show()

# Median value
median_projects = df['numberOfProjects'].median()
print("Median number of projects:", median_projects)
```

5. Plot a histogram using the variable “avgMonthlyHours” and find the range in which the number of employees worked for 150 hours per month?

```
# Histogram for avgMonthlyHours
plt.figure(figsize=(8, 6))
plt.hist(df['avgMonthlyHours'], bins=20, edgecolor='black')
plt.title('Histogram of avgMonthlyHours')
plt.xlabel('Average Monthly Hours')
plt.ylabel('Frequency')
plt.show()

# Range for 150 hours
num_employees_150_hours = ((df['avgMonthlyHours'] >= 150) & (df['avgMonthlyHours'] < 160)).sum()
print("Number of employees worked 150 hours per month:", num_employees_150_hours)
```

6. Generate a boxplot for the variables “lastEvaluation” and “numberOfProjects”

```
import seaborn as sns

# Boxplot for lastEvaluation and numberOfProjects
plt.figure(figsize=(10, 6))
sns.boxplot(data=df[['lastEvaluation', 'numberOfProjects']])
plt.title('Boxplot of lastEvaluation and numberOfProjects')
plt.ylabel('Value')
plt.show()
```