

Step 1: EDA Project 1 - Analysis of AMCAT Data - Introduction

Data Description: The dataset used for this analysis is derived from Aspiring Minds Employment Outcome 2015 (AMEO). It contains the employment outcomes of engineering graduates, primarily focusing on their salary, job titles, and job locations. The dataset consists of approximately 4000 data points and 40 independent variables, which are a mix of continuous and categorical features.

These variables encompass a wide range of information, such as:

Demographic details: Gender, Date of Birth (DOB), Job Location, etc. **Educational background:** 10th and 12th percentages, college GPA, specialization, degree, etc. **Technical skills:** Scores in domain-specific sections such as Computer Programming, Electrical Engineering, etc. **Personality traits:** Scores in traits like conscientiousness, agreeableness, extraversion, neuroticism, and openness to experience. The dataset also includes two important date fields: the Date of Joining (DOJ) and Date of Leaving (DOL), along with unique candidate identifiers.

Objective: The primary goal of this exploratory data analysis (EDA) is to understand the factors that influence the Salary offered to the engineering graduates. The analysis will focus on the following objectives:

Understand the overall data structure: Check for missing values, data types, and distributions of various variables. Explore the distribution of salary: Analyze how salaries vary across different features such as gender, specialization, technical scores, and personality traits. Identify key relationships: Investigate correlations between salary and other features such as education, job location, and skills. Discover trends and insights: Uncover patterns in the data that can be helpful in understanding employment outcomes in terms of salary. The analysis will follow standard EDA steps to gain insights and potentially guide decision-making for future analyses or predictive modeling.

```
[12]: # Step 2 - Import necessary libraries
import pandas as pd

# Step 2 - Load the dataset
# Assuming the dataset is in csv format; replace 'data.csv' with the actual file path if different
df = pd.read_excel('data.xlsx')

# Step 2 - Display the first few rows of the dataset (head)
print("Data Head:")
print(df.head())

# Step 2 - Display the shape of the dataset
print("\nShape of the dataset (rows, columns):")
print(df.shape)

# Step 2 - Display basic statistics and descriptions of the dataset
print("\nSummary Statistics (for continuous variables):")
print(df.describe())

# Step 2 - Display information about the dataset (data types, non-null counts)
print("\nDataset Information:")
print(df.info())
```

	Unnamed: 0	ID	Salary	DOJ	DOL	\
0	train	203097	4200000	2012-06-01	present	
1	train	579905	5000000	2013-09-01	present	
2	train	810601	3250000	2014-06-01	present	
3	train	267447	11000000	2011-07-01	present	
4	train	343523	2000000	2014-03-01	2015-03-01 00:00:00	



jupyter project1_innomatics NEW Last Checkpoint 29 minutes ago

File Edit View Run Kernel Settings Help Not Trusted

Step 3: Univariate Analysis

```
[13]: # Import Necessary Libraries: Make sure to import the libraries required for plotting and analysis.
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

[14]: # Set Up Visualization Styles: Set up the visualization styles for better aesthetics
sns.set(style='whitegrid')

[15]: # Identify Numerical and Categorical Columns: Extract numerical and categorical columns from the dataset.
numerical_cols = df.select_dtypes(include=['float64', 'int64']).columns.tolist()
categorical_cols = df.select_dtypes(include=['object']).columns.tolist()

[17]: # Univariate Analysis for Numerical Variables: Let's create histograms, box plots, and identify outliers for each numerical column.
for col in numerical_cols:
    plt.figure(figsize=(14, 6))

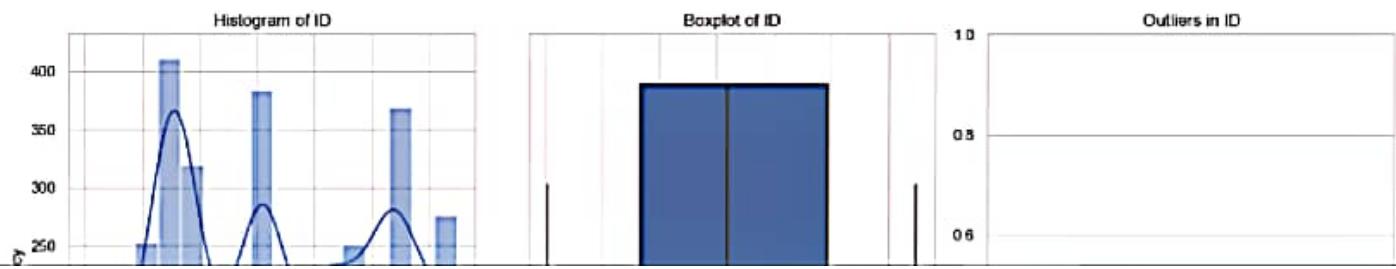
    # Histogram
    plt.subplot(1, 3, 1)
    sns.histplot(df[col], kde=True)
    plt.title(f'Histogram of {col}')
    plt.xlabel(col)
    plt.ylabel('Frequency')

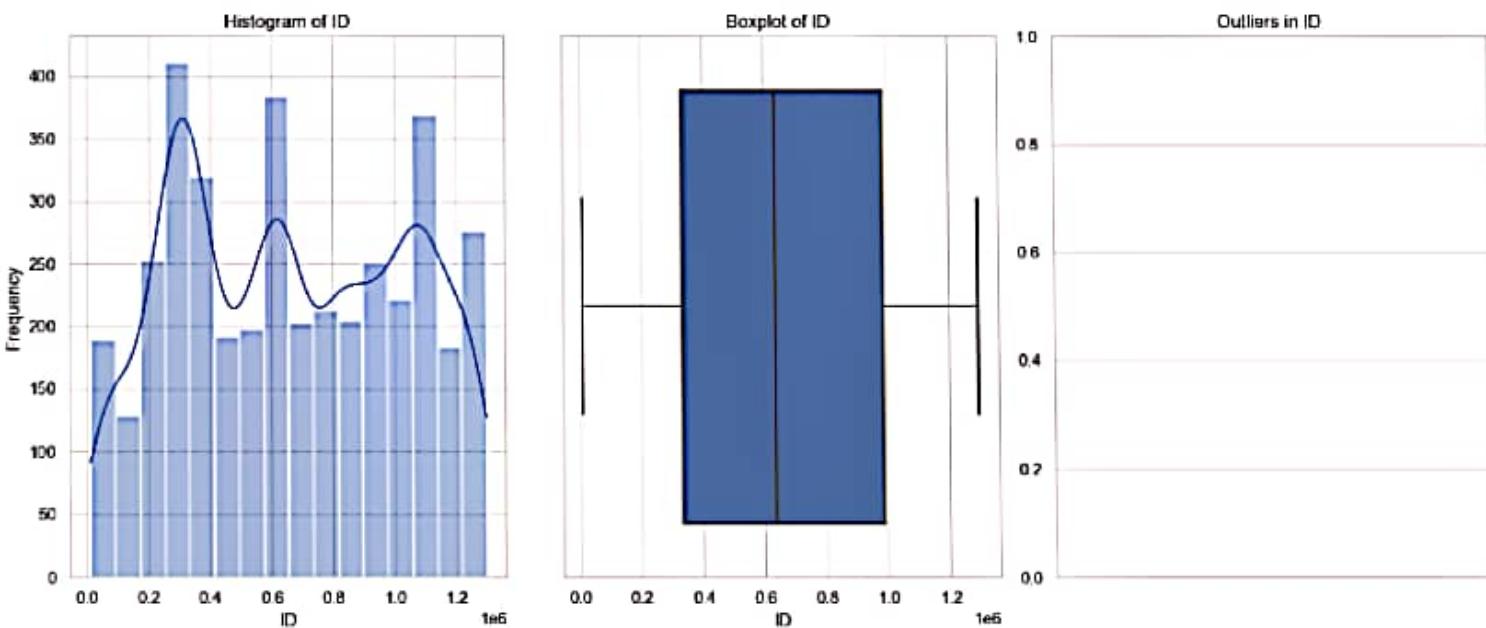
    # Boxplot
    plt.subplot(1, 3, 2)
    sns.boxplot(x=df[col])
    plt.title(f'Boxplot of {col}')

    # Outliers
    outliers = df[(df[col] < df[col].quantile(0.25) - 1.5 * (df[col].quantile(0.75) - df[col].quantile(0.25))) | (df[col] > df[col].quantile(0.75) + 1.5 * (df[col].quantile(0.75) - df[col].quantile(0.25)))]
    plt.subplot(1, 3, 3)
    sns.scatterplot(x=[1]*len(outliers), y=outliers[col], color='red')
    plt.title(f'Outliers in {col}')
    plt.xticks([])

    plt.tight_layout()
    plt.show()

    print(f'Outliers in {col}:\n', outliers)
```

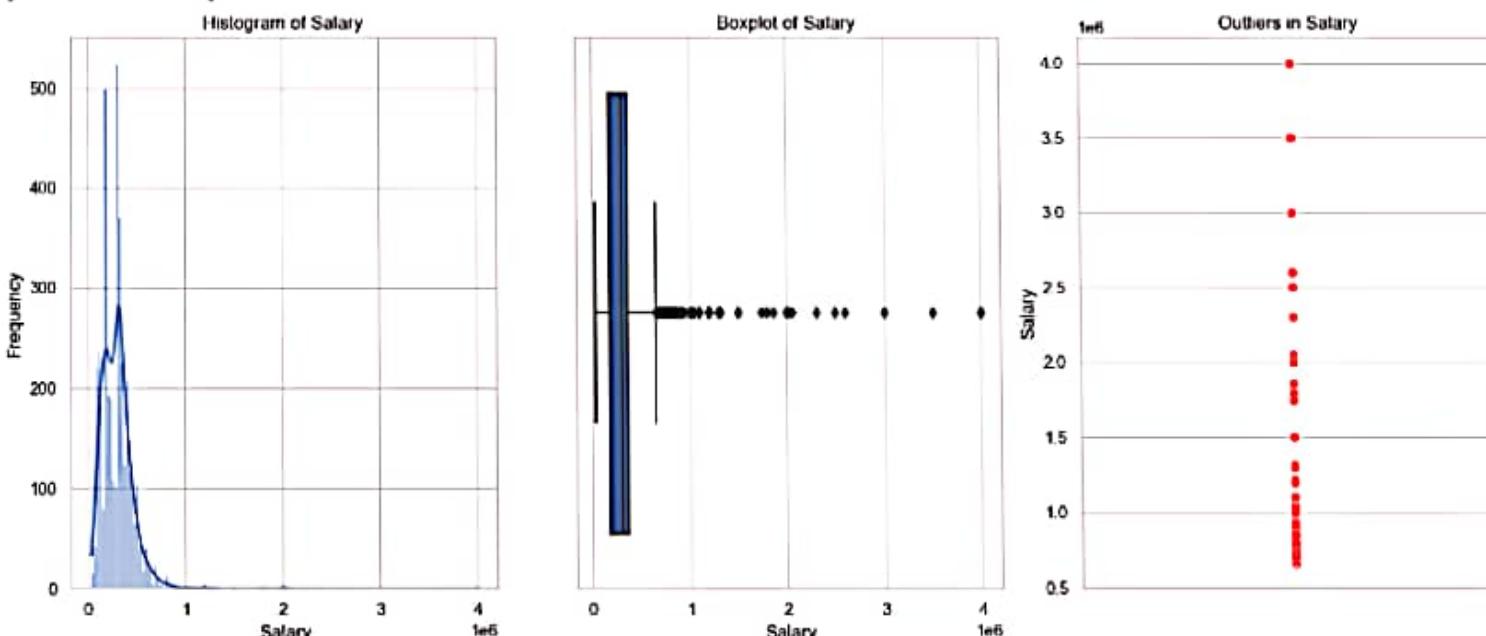




Outliers in ID:

```
Empty DataFrame
Columns: [Unnamed: 0, ID, Salary, DOJ, DOL, Designation, JobCity, Gender, DOB, 10percentage, 10board, 12graduation, 12percentage, 12board, CollegeID, CollegeTier, Degree, Specialization, collegeGPA, CollegeCityID, CollegeCityTier, CollegeState, GraduationYear, English, Logical, Quant, Domain, ComputerProgramming, ElectronicsAndSemicon, ComputerScience, MechanicalEngg, ElectricalEngg, TelecomEngg, CivilEngg, conscientiousness, agreeableness, extraversion, neuroticism, openness_to_experience]
Index: []
```

[0 rows x 39 columns]



Scanned with OKEN Scanner

```

Unnamed: 0   ID    Salary      DOJ          DOL \
3     train  267447 1100000 2011-07-01      present
76    train  361583 800000 2012-06-01      present
92    train  1250429 1500000 2014-11-01 2014-07-01 00:00:00
123   train  312164 1200000 2010-07-01 2011-07-01 00:00:00
128   train  206734 675000 2011-11-01      present
...
...
3823  train  918568 775000 2014-08-01      present
3984  train  267121 850000 2011-09-01      present
3912  train  231229 730000 2013-07-01      present
3961  train  230782 700000 2011-07-01 2014-09-01 00:00:00
3992  train  344407 800000 2014-04-01 2015-04-01 00:00:00

      Designation      JobCity Gender      DOB 10percentage \
3   senior software engineer      Gurgaon   m 1989-12-05    85.60
76   software engineer      Bangalore   m 1991-01-25    93.44
92   application developer      Hyderabad   m 1992-01-04    79.00
123  engineer trainee      Maharaigangj   m 1988-04-25    59.80
128  senior software engineer      Noida   m 1988-11-07    68.00
...
...
3823 mechanical design engineer      Dammam   m 1991-01-12    87.48
3984   operations assistant      Noida   m 1989-01-05    83.40
3912   research scientist      Pune   m 1989-11-15    84.67
3961   planning engineer      Rajpura   m 1987-12-27    84.20
3992       manager      Rajkot   m 1990-06-22    73.00

... ComputerScience MechanicalEngg ElectricalEngg TelecomEngg \
3     ...      -1      -1      -1      -1
76     ...      -1      -1      -1      -1
92     ...      346      -1      -1      -1
123    ...      -1      206      -1      -1
128    ...      -1      -1      -1      -1
...
...
3823   ...      -1      469      -1      -1
3984   ...      -1      -1      -1      -1
3912   ...      -1      -1      -1      -1
3961   ...      -1      -1      -1      -1
3992   ...      -1      -1      -1      -1

      CivilEngg conscientiousness agreeableness extraversion neuroticism \
3        -1      0.0464      0.3448     -0.3448     -0.40780
76        -1     -0.4173      0.9688     -0.1988     -0.29820
92        -1      0.4155      0.5454      0.9322     -0.61470
123        -1      0.2009      1.1248      1.1074     -1.11280
128        -1     -0.8810     -0.2793     -0.6343     -0.64280
...
...
3823     -1     -0.8772     -0.1206     -0.1437     -0.23440
3984     -1     -0.8810      0.1888     -0.1988     -0.85520
3912     -1     -1.3447     -1.0593      0.6720      1.00240
3961     460     -1.3447      0.8328     -2.3759     -0.99530
3992     480      0.3555     -0.9033      0.9623      0.64983

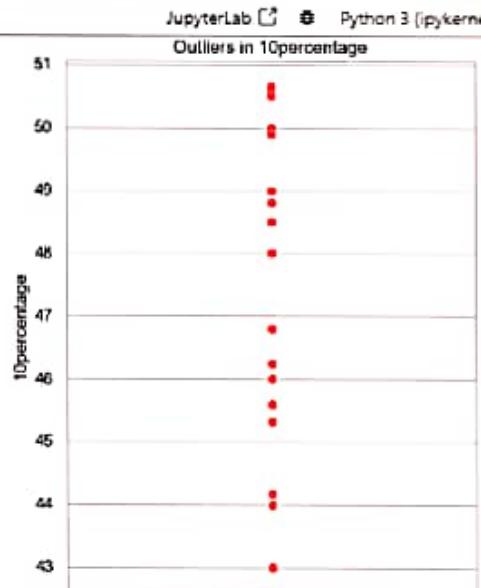
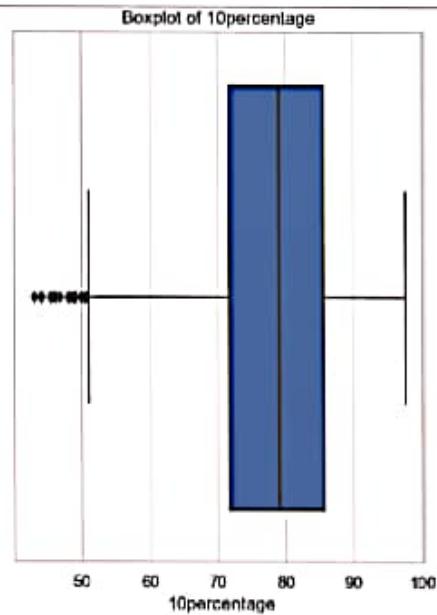
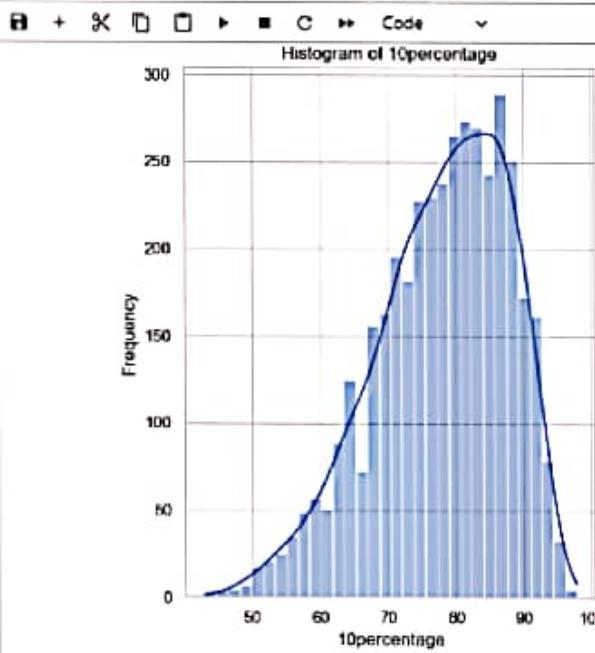
      openness_to_experience
3           -0.9194
76            0.3849
92            0.8637
123            0.9763
128            2.9731
...
...
3823            -0.8943
3984            -1.0774
3912            -1.7893
3961            0.3444

```

jupyter project1_innomatics NEW Last Checkpoint: 29 minutes ago

File Edit View Run Kernel Settings Help

Not Tru



Outliers in 10percentage:

	Unnamed: 0	ID	Salary	DOJ	DOL	\
245	train	984958	285000	2014-08-01	2015-04-01	00:00:00
466	train	1245184	215000	2015-01-01	2015-04-01	00:00:00
490	train	382937	150000	2011-07-01	2012-11-01	00:00:00
491	train	53838	100000	2010-05-01	2011-01-01	00:00:00
582	train	1160766	240000	2015-01-01	2015-04-01	00:00:00
680	train	918414	200000	2008-11-01	2012-11-01	00:00:00
613	train	228881	440000	2013-05-01		present
898	train	212005	270000	2013-03-01		present
919	train	1231953	450000	2014-07-01	2015-06-01	00:00:00
1864	train	277260	250000	2011-09-01	2014-12-01	00:00:00
1182	train	311098	120000	2012-12-01	2015-01-01	00:00:00
1169	train	179878	105000	2011-01-01	2011-06-01	00:00:00
1193	train	309190	180000	2012-12-01	2013-12-01	00:00:00
1235	train	325701	115000	2012-11-01	2013-01-01	00:00:00
1334	train	240133	110000	2012-01-01	2012-08-01	00:00:00
1838	train	637345	190000	2014-01-01	2015-04-01	00:00:00
1845	train	1251526	150000	2014-12-01		present
1955	train	601296	145000	2013-08-01	2014-08-01	00:00:00
1976	train	636483	105000	2013-05-01	2014-12-01	00:00:00
2817	train	985445	120000	2014-11-01	2015-04-01	00:00:00
2215	train	10054880	235000	2013-09-01	2014-06-01	00:00:00
2292	train	457936	300000	2013-03-01	2014-09-01	00:00:00
2432	train	878844	120000	2014-06-01	2015-01-01	00:00:00
2655	train	1089862	85000	2012-08-01	2011-05-01	00:00:00
2885	train	505714	240000	2014-01-01	2015-02-01	00:00:00
2982	train	1251571	180000	2014-10-01	2015-05-01	00:00:00
3284	train	267354	120000	2012-02-01	2013-10-01	00:00:00
3425	train	1269533	120000	2014-01-01	2014-05-01	00:00:00
3690	train	812583	290000	2013-01-01		present
3743	train	63887	395000	2011-11-01		present
	Designation		JobCity	Gender	DOB	\
245	technical support engineer		Jaipur		= 1991-10-05	
466	software developer		New Delhi		= 1989-10-10	
490	design engineer		Pune		= 1989-05-05	

jupyter project1_innomatics NEW Last Checkpoint: 29 minutes ago



File Edit View Run Kernel Settings Help

Not Trusted

JupyterLab ⚙ Python 3 (ipykernel)

```
245 technical support engineer Jaipur m 1991-10-05
466 software developer New Delhi m 1989-10-10
498 design engineer Pune m 1989-05-05
491 software developer Bhubaneshwar m 1987-06-12
502 electrical field engineer New Delhi m 1990-11-04
608 project engineer Thiruvananthapuram m 1986-05-27
613 operations manager Missar m 1988-10-25
898 asp.net developer Noida m 1989-01-24
919 system engineer Bangalore f 1989-07-12
1064 .net developer New Delhi m 1987-12-28
1182 .net developer New Delhi m 1983-09-02
1169 software trainee New Delhi m 1985-02-15
1193 network engineer Noida m 1986-02-06
1235 phone banking officer Hyderabad f 1989-07-24
1334 electrical engineer ncr m 1987-08-06
1838 quality analyst Hyderabad m 1991-07-03
1845 office coordinator New Delhi f 1993-11-23
1955 .net developer -1 m 1989-06-23
1976 software developer Bangalore m 1989-11-21
2837 technical support executive A-64,sec-64,noida m 1991-06-15
2215 sales executive Noida m 1992-03-15
2292 project engineer Bangalore m 1998-09-01
2412 java software engineer Noida m 1993-01-18
2655 software engineer Pune m 1987-01-21
2885 project coordinator Jaipur m 1989-10-19
2982 service coordinator Noida m 1991-06-04
3284 quality analyst Noida m 1988-08-05
3425 web developer Pune m 1990-02-25
3600 asp.net developer Bangalore m 1985-01-15
3743 senior software engineer -1 m 1985-05-11
```

```
10percentage ... ComputerScience MechanicalEngg ElectricalEngg \
```

```
245 50.60 ... 315 -1 -1
466 44.16 ... -1 -1 -1
498 44.00 ... -1 -1 -1
491 45.60 ... -1 -1 -1
502 48.00 ... -1 -1 516
608 49.00 ... -1 -1 -1
613 48.00 ... -1 -1 -1
898 49.00 ... -1 -1 -1
919 48.00 ... 500 -1 -1
1064 49.00 ... -1 -1 -1
1182 49.00 ... -1 -1 -1
1169 48.50 ... -1 -1 -1
1193 48.00 ... -1 -1 -1
1235 50.60 ... -1 -1 -1
1334 43.00 ... -1 -1 -1
1838 50.00 ... 253 -1 -1
1845 49.00 ... -1 -1 -1
1955 45.33 ... -1 -1 -1
1976 46.24 ... -1 -1 -1
2837 48.00 ... 346 -1 -1
2215 50.50 ... 487 -1 -1
2292 50.00 ... -1 -1 -1
2412 50.00 ... 438 -1 -1
2655 50.66 ... -1 -1 -1
2885 46.80 ... -1 -1 -1
2982 50.00 ... 284 -1 -1
3284 50.00 ... -1 -1 -1
3425 50.00 ... -1 -1 -1
3600 46.00 ... -1 -1 -1
3743 49.00 ... -1 -1 -1
```

	TelecomEngg	CivilEngg	conscientiousness	agreeableness	extraversion	\
245	-1	-1	-0.0154	0.2124	-0.1437	
466	-1	-1	0.2718	0.8784	0.8100	
498	286	-1	-0.5180	0.0328	0.6984	
491	-1	-1	0.6646	0.3448	-1.0697	
502	-1	-1	-0.5236	-1.1196	0.1637	
600	-1	-1	-1.1644	-1.9521	-1.0659	
613	-1	-1	0.0464	0.6568	-0.0537	
898	-1	-1	-0.8818	-1.5273	-0.3448	
919	-1	-1	0.9900	0.3789	1.0859	
1064	-1	-1	0.8192	0.1888	-0.3440	
1102	-1	-1	1.2828	0.8128	0.6728	
1169	-1	-1	-2.1175	-3.0874	-2.6662	
1193	313	-1	-0.1882	-0.1232	0.3817	
1235	-1	-1	-0.4173	0.1888	-0.1988	
1334	-1	-1	0.9737	0.6568	0.5259	
1838	286	-1	-0.1590	-0.1206	0.3174	
1845	292	-1	-0.3027	1.2114	0.1637	
1955	-1	-1	-1.7389	-0.2871	-0.7026	
1976	-1	-1	0.8463	0.3789	0.6248	
2037	-1	-1	-1.3080	-1.7856	0.0100	
2215	-1	-1	0.7827	0.0459	0.3174	
2292	-1	-1	0.5100	0.0328	-0.0537	
2432	-1	-1	0.4155	-0.6584	0.3034	
2655	-1	-1	1.2828	-1.3713	0.8171	
2885	-1	-1	-1.3447	-0.2793	-0.4891	
2982	-1	-1	1.5644	0.8784	0.6248	
3284	-1	-1	-0.4173	-1.3713	0.9842	
3425	-1	-1	-0.0154	-0.1206	-0.2974	
3690	-1	-1	0.2718	-0.1206	-0.6848	
3743	-1	-1	0.8192	1.1248	0.6728	

neuroticism openness_to_experience

245	-1.24858	1.2478
456	-0.18760	0.4805
498	-0.56448	0.6603
491	0.29738	-0.9194
502	0.01920	-1.0872
600	0.52620	-1.4356
613	-0.05520	0.6603
898	0.53230	-0.9194
919	0.65300	0.4805
1064	-0.76030	0.8183
1102	-1.70040	-0.2875
1169	0.06230	-3.4471
1193	-0.48780	-1.0774
1235	0.88483	-0.1295
1334	0.88480	-0.1295
1838	0.01920	0.2889
1845	0.39950	1.2470
1955	0.52620	-1.1291
1976	1.28690	0.4805
2037	0.81920	-2.3937
2215	0.27270	0.2889
2292	0.88480	0.8183
2432	0.65300	-0.4392
2655	0.88480	0.5824
2885	1.50180	-0.9194
2982	0.98660	0.8637
3284	1.55780	0.3444
3425	-0.61470	-0.4776
3690	-0.86820	-0.2859
3743	0.86230	1.1343



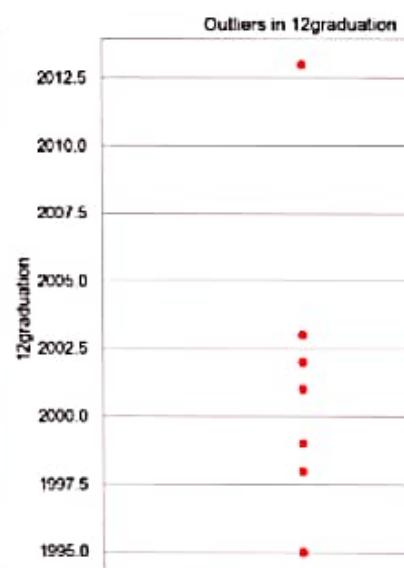
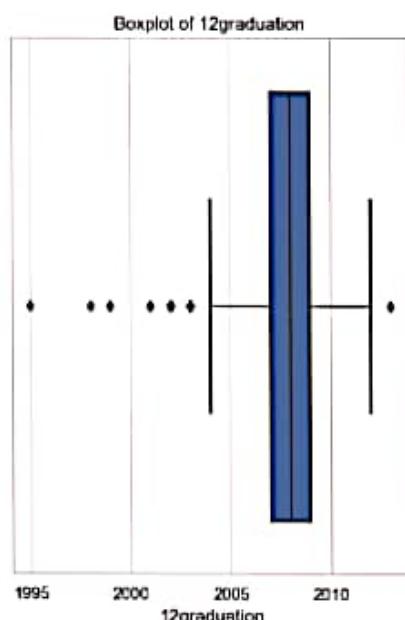
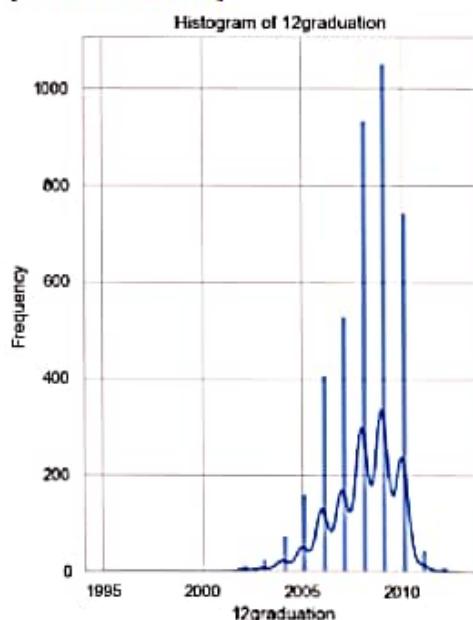
Project 1: Data Cleaning and Analysis

jupyter project1_innomatics NEW Last Checkpoint: 29 minutes ago

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

[38 rows x 39 columns]



Outliers in 12graduation:

	Unnamed: 0	ID	Salary	DOJ	DOL	\
59	train	536853	1280000	2809-09-01	2013-04-01 00:00:00	
193	train	231253	550000	2811-08-01	2015-04-01 00:00:00	
528	train	289374	250000	2811-05-01	2011-09-01 00:00:00	
584	train	283323	430000	2010-04-01	present	
600	train	918414	200000	2008-11-01	2012-11-01 00:00:00	
670	train	50670	150000	2010-02-01	2013-02-01 00:00:00	
712	train	114364	400000	2010-11-01	2013-09-01 00:00:00	
772	train	1080456	320000	2013-07-01	2015-02-01 00:00:00	
828	train	54579	195000	2010-04-01	2011-12-01 00:00:00	
898	train	134985	310000	2010-09-01	2015-02-01 00:00:00	
1083	train	243897	400000	2012-01-01	present	
1169	train	179878	105000	2811-01-01	2011-06-01 00:00:00	
1178	train	266142	320000	2811-05-01	2013-09-01 00:00:00	
1484	train	626050	700000	2813-06-01	present	
1420	train	59151	200000	2010-01-01	2013-01-01 00:00:00	
1582	train	678819	200000	2813-05-01	2013-07-01 00:00:00	
1528	train	257095	110000	2010-08-01	2011-10-01 00:00:00	
1866	train	187787	150000	2010-10-01	2012-03-01 00:00:00	
1873	train	635518	325000	2014-03-01	2015-04-01 00:00:00	
1883	train	36817	100000	2811-05-01	2012-04-01 00:00:00	
1927	train	574430	180000	2811-10-01	2013-04-01 00:00:00	
1982	train	232481	2500000	2012-04-01	present	
2009	train	40772	650000	2010-07-01	2013-02-01 00:00:00	
2217	train	249696	180000	2012-10-01	2014-09-01 00:00:00	
2358	train	78948	420000	2010-10-01	2014-02-01 00:00:00	
2409	train	48796	680000	2010-05-01	2014-11-01 00:00:00	
2506	train	56681	380000	2010-02-01	2013-05-01 00:00:00	
2674	train	78728	240000	2010-07-01	2011-01-01 00:00:00	
2746	train	53545	330000	2010-11-01	2012-02-01 00:00:00	
2747	train	78777	750000	2010-09-01	2014-01-01 00:00:00	
3015	train	38335	720000	2810-07-01	present	
3193	train	488729	12NNNN	2007-06-01	2009-06-01 00:00:00	

jupyter project1_innomatics NEW Last Checkpoint 29 minutes ago

File Edit View Run Kernel Settings Help

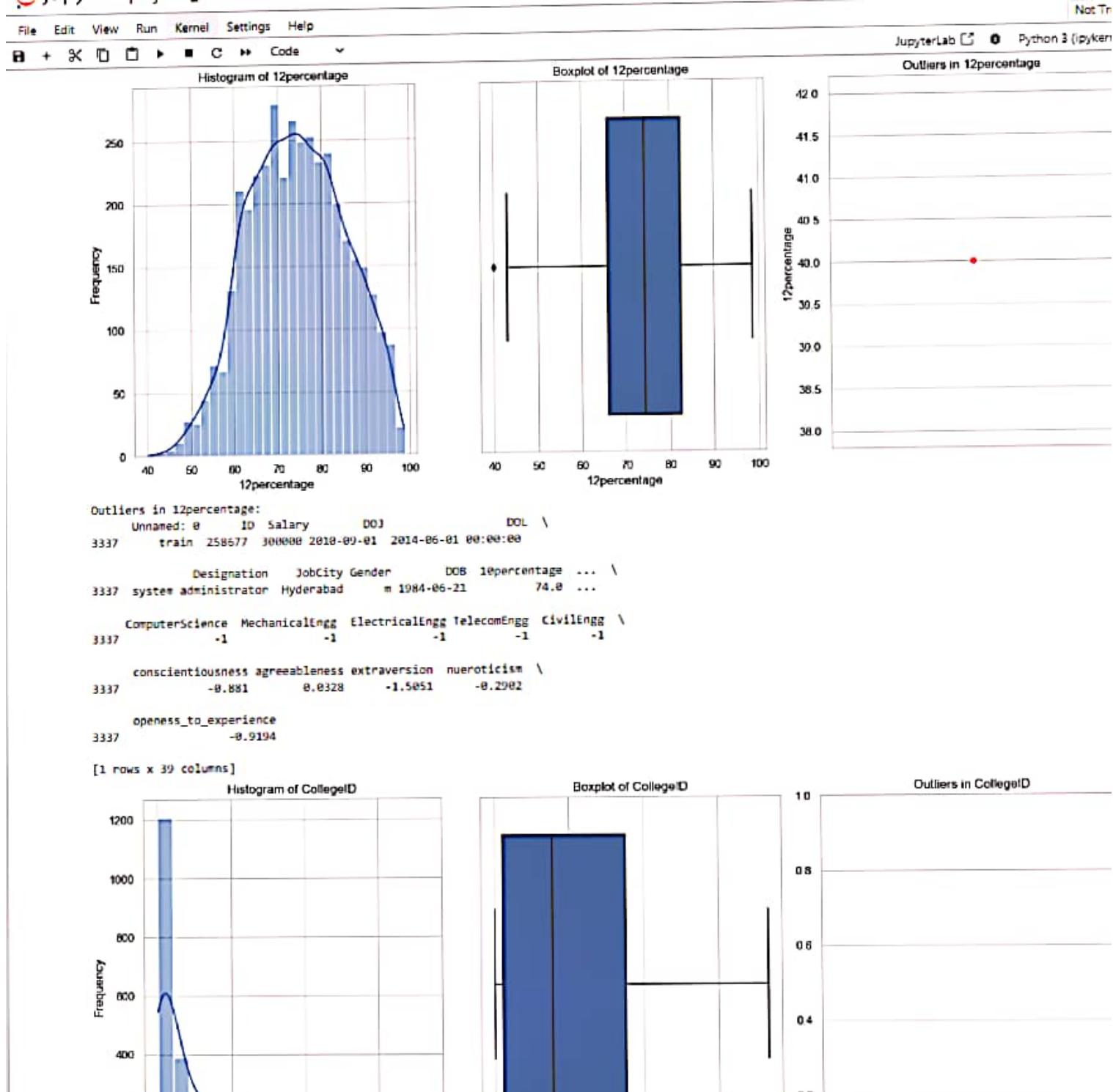
Not Trusted

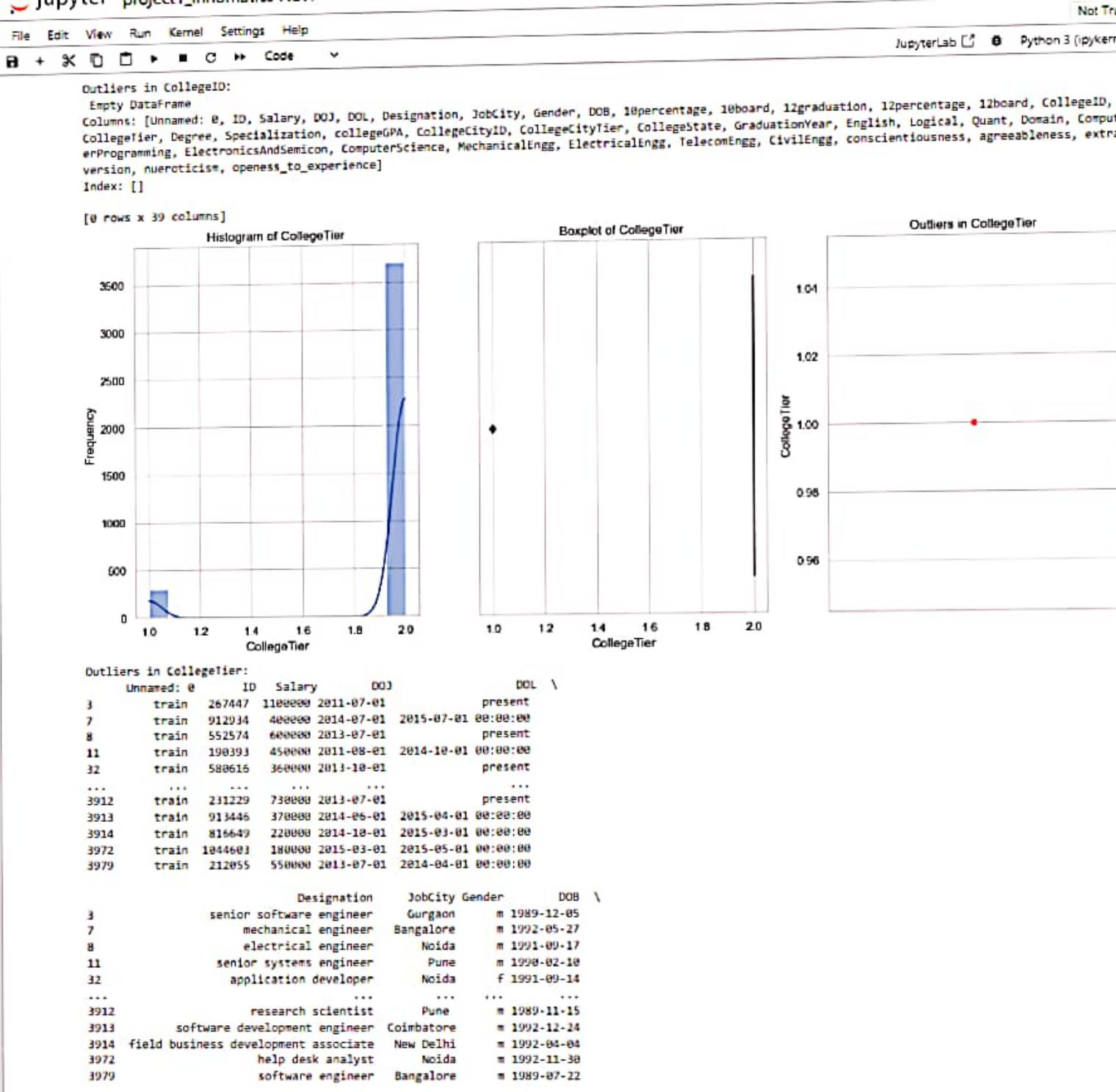
JupyterLab Python 3 (ipykernel)

	Designation	JobCity	Gender	DOB
59	software engineer	Bangalore	m	1977-10-30
193	software developer	Mumbai	m	1985-07-19
528	design engineer	New Delhi	m	1986-05-29
584	lead engineer	Bangalore	m	1986-11-03
608	project engineer	Thiruvananthapuram	m	1986-05-27
678	project engineer	Bangalore	f	1985-11-03
712	programmer analyst	Kochi/Cochin	m	1985-02-08
772	java developer	Baroda	m	1985-04-18
828	technical recruiter	-1	m	1985-07-01
898	software developer	Kolkata	m	1983-10-23
1083	assistant professor	New Delhi	f	1983-02-23
1169	software trainee	New Delhi	m	1985-02-15
1178	software engineer	Hyderabad	m	1986-05-10
1484	senior software engineer	Noida	m	1985-05-16
1420	administrative support	Bhubaneswar	f	1985-06-01
1502	sales executive	New Delhi	m	1983-01-10
1528	desktop support engineer	Chennai	m	1981-05-06
1866	java software engineer	Ahmedabad	m	1985-05-02
1873	software engineer	Pune	f	1991-11-25
1883	seo executive	Noida	m	1985-01-03
1927	software engineer	Hyderabad	m	1986-10-28
1982	software test engineer	-1	m	1984-09-21
2009	assistant manager	Noida	m	1984-09-03
2117	.net developer	Kolkata	m	1985-05-16
2358	software engineer	Noida	m	1985-07-01
2489	software engineer	Bangalore	m	1984-07-25
2606	software engineer	Bangalore	m	1982-07-22
2674	rf engineer	-1	m	1986-02-21
2746	software engineer	Pune	m	1984-10-29
2747	senior software engineer	Bangalore	m	1987-10-25
3015	civil engineer	Patna	m	1985-11-17
3193	software engineer	Hyderabad	m	1985-06-18
3237	software engineer	Bangalore	m	1985-04-08
3278	software developer	New Delhi	m	1982-01-25
3292	test engineer	Bangalore	m	1984-11-26
3308	engineer	Aurangabad	m	1985-06-07
3322	senior software developer	-1	m	1986-11-01
3337	system administrator	Hyderabad	m	1984-06-21
3431	field engineer	Muzaffarpur	m	1987-01-18
3516	programmer	Hyderabad	f	1985-05-15
3655	application developer	New Delhi	m	1985-02-23
3690	asp.net developer	Bangalore	m	1985-01-15
3743	senior software engineer	-1	m	1985-05-11
3796	web developer	Greater Noida	m	1985-04-14
3921	data analyst	Bangalore	m	1984-11-19

	10percentage	ComputerScience	MechanicalEngg	ElectricalEngg
59	72.00	...	-1	366
193	82.20	...	-1	-1
528	77.00	...	-1	473
584	61.00	...	-1	-1
608	49.00	...	-1	-1
678	75.84	...	-1	-1
712	91.00	...	-1	-1
772	65.80	...	487	-1
828	64.83	...	-1	-1
898	75.00	...	-1	-1
1083	70.00	...	-1	-1
1169	48.50	...	-1	-1
1178	67.67	...	-1	-1
1404	75.60	...	-1	-1

	TelecomEngg	CivilEngg	conscientiousness	agreeableness	extraversion	\
59	-1	400	-0.3814	0.4353	-0.9245	
193	206	-1	-1.3447	0.6568	0.3817	
528	-1	-1	-0.4173	-0.5913	0.6914	
584	-1	-1	-0.8810	0.6568	0.6720	
600	-1	-1	-1.1644	-1.9521	-1.0659	
670	-1	-1	0.3555	-0.7473	-0.1988	
712	-1	-1	0.8192	1.4368	-0.8537	
772	-1	-1	-0.7335	-0.4536	0.4711	
828	-1	-1	0.2009	0.5008	0.8171	
890	-1	-1	-0.4173	-0.1232	-0.0537	
1083	393	-1	-1.8083	0.3448	-1.3599	
1169	-1	-1	-2.1175	-3.0874	-2.6662	
1178	-1	-1	-0.1082	1.1248	0.9623	
1404	-1	-1	0.9900	1.2114	0.6248	
1420	-1	-1	0.9737	0.9688	1.5428	
1582	-1	-1	0.5591	1.5444	-0.1437	
1528	-1	-1	-1.3447	-1.2153	1.1074	
1866	-1	-1	0.5100	1.2808	0.8171	
1873	-1	-1	1.2772	0.7119	0.1537	
1883	-1	-1	-0.4173	-2.1513	0.5269	
1927	-1	-1	1.4374	-0.5013	1.1074	
1982	-1	-1	0.6646	0.8128	0.3817	
2009	-1	-1	-0.2528	-0.2793	-0.3448	
2217	-1	-1	0.2009	-1.2153	-0.7794	
2358	-1	-1	0.2009	0.0328	1.2525	
2409	-1	-1	0.0464	0.5008	-0.0537	
2506	-1	-1	-0.7264	0.1888	-0.7794	
2674	-1	-1	-0.1082	1.4368	1.2525	
2746	-1	-1	-0.2628	0.8328	-1.5051	
2747	-1	-1	-0.5719	1.2808	0.2366	
3015	-1	300	0.6646	-0.1232	0.2366	
3193	-1	-1	-1.4992	-0.7863	-2.0131	
3237	-1	-1	0.5100	0.9688	1.6880	
3278	-1	-1	-1.6538	-0.9033	-0.7794	
3292	-1	-1	-0.5719	-0.1232	-0.0537	
3308	420	-1	-1.3447	1.1248	0.6720	
3322	-1	-1	0.0464	0.9688	0.3817	
3337	-1	-1	-0.8810	0.0328	-1.5051	
3431	-1	-1	0.6646	-0.2793	-0.0537	
3516	-1	-1	0.8192	0.2668	-0.5988	
3655	-1	-1	0.0464	0.6568	0.3817	
3698	-1	-1	0.2718	-0.1206	-0.6048	
3743	-1	-1	0.8192	1.1248	0.6720	
3796	-1	-1	-0.4595	0.0328	-1.7083	
3921	-1	-1	0.0464	0.2668	-1.7954	
			maseroticism	openness_to_experience		
59	-0.34900		-0.4455			
193	-0.64280		-0.1295			
528	-0.99530		-0.6835			
584	1.47240		0.3444			
600	0.52620		-1.4356			
670	0.76730		-0.4455			
712	-1.81700		0.8183			
772	0.27270		0.4805			
828	0.41480		0.3444			
890	-0.48780		-0.1295			
1083	1.58090		1.2923			
1169	0.06230		-3.4471			
1178	-0.99530		0.1864			
1404	-1.37530		0.6721			
			*	*	*	





 jupyter project1_innomatics NEW Last Checkpoint: 30 minutes ago


Not Trusted

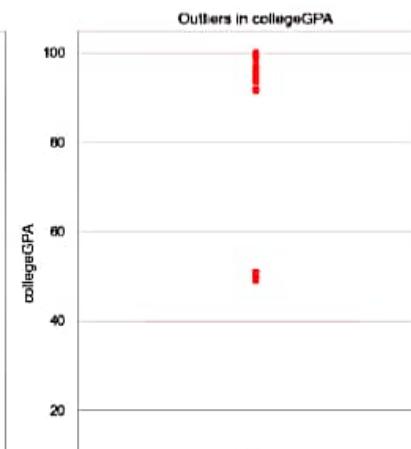
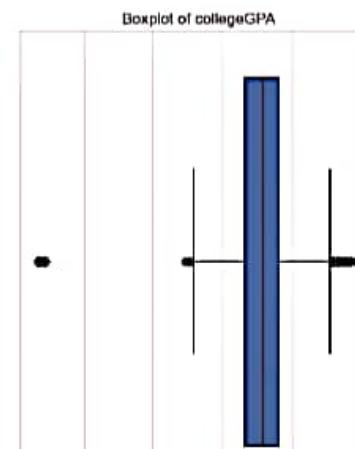
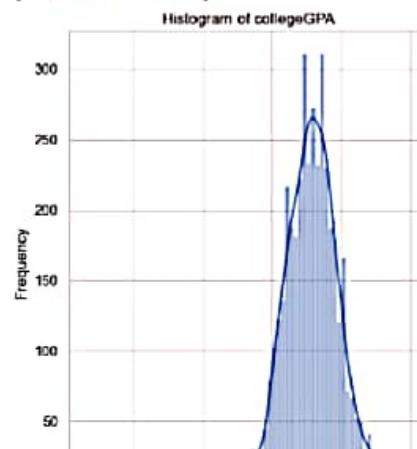
JupyterLab Python 3 (ipykernel)

```
10percentage ... ComputerScience MechanicalEngg ElectricalEngg \
3      85.60 ... -1       -1       -1
7      92.00 ... -1       469      -1
8      90.00 ... -1       -1       -1
11     81.00 ... -1       -1       -1
32     87.00 ... -1       -1       -1
...     ... ...
3912    84.67 ... -1       -1       -1
3913    77.80 ... 438      -1       -1
3914    71.80 ... -1       -1       -1
3972    68.20 ... 487      -1       -1
3979    69.16 ... -1       -1       -1
```

```
TelecomEngg civilEngg conscientiousness agreeableness extraversion \
3      -1       -1       0.0464   0.3448   -0.3448
7      -1       -1      -0.0154   1.2114   0.0100
8      -1       -1      -0.1598   0.5454   -0.6048
11     -1       -1      -2.2720   1.1248   0.8171
32     -1       -1      -0.7335   -0.4536  -1.0659
...     ... ...
3912    -1      -1      -1.3447  -1.0593   0.6720
3913    -1      -1      0.2718   0.0459   0.1637
3914    260     -1      -0.7335  -0.6201   0.6248
3972    -1      -1      -1.1544   0.2124   -0.2974
3979    -1      -1      -0.5719   0.5088   -0.4891
```

```
neuroticism openness_to_experience
3      -0.4078   -0.9194
7      0.1459   1.2478
8      -0.7415   -0.2859
11     -0.4078   -0.4455
32     1.1601   -0.4776
...     ...
3912    1.0024   -1.7093
3913    1.4116   -2.2021
3914    1.5404   -0.6692
3972    0.8192   -1.0524
3979    0.4148   -1.2354
```

[297 rows x 39 columns]



	TelecomEngg	CivilEngg	conscientiousness	agreeableness	extraversion
7	-1	-1	-0.8154	1.2114	0.8188
44	-1	-1	0.5188	0.6568	-0.6343
138	-1	-1	0.8463	0.7119	0.4711
187	-1	-1	0.5188	1.1248	0.5269
477	393	-1	-0.6491	-0.4353	0.6728
614	348	-1	-0.4173	0.6568	-0.9245
698	313	-1	0.8463	-0.6281	-0.7585
788	-1	-1	0.5188	-0.2793	-1.0697
874	-1	-1	0.5188	0.9688	0.8171
987	-1	-1	-0.5332	-0.4353	-0.1988
1829	-1	-1	-0.5719	1.0858	-0.9245
1134	-1	-1	1.5928	0.9588	0.9623
1264	-1	-1	-1.6538	-0.5913	0.2366
1345	-1	-1	-0.3027	0.8459	-1.6887
1419	-1	-1	1.8517	0.2124	0.6248
1439	-1	260	-2.7357	-2.3873	-0.7794
1518	-1	-1	-0.8696	0.6568	-0.6343
1685	446	-1	0.2718	-0.6281	-0.6048
1767	-1	-1	1.9953	-0.9531	-1.6887
2151	-1	-1	-1.0355	0.6568	0.9623
2152	-1	-1	1.4374	1.7488	0.9623
2229	-1	-1	0.4155	-0.6281	0.8188
2293	-1	-1	1.1283	1.4368	-0.7058
2463	-1	-1	0.6646	1.7488	0.3817
2662	-1	-1	-0.4173	-1.6833	0.5269
2691	-1	-1	-0.8818	-0.8842	-0.4891
2783	-1	-1	0.5591	0.5454	-0.6048
2836	-1	-1	-1.2287	-0.2793	-0.4891
2880	-1	-1	-0.1598	0.5454	0.4711
2988	-1	-1	0.1282	0.2124	0.1637
3151	-1	-1	0.9737	0.6568	0.2366
3276	-1	-1	-1.4992	0.3448	-0.8537
3293	-1	-1	-0.4173	0.6568	0.8171
3388	428	-1	-1.3447	1.1248	0.6728
3323	393	-1	1.5644	0.7119	0.4711
3448	393	-1	-2.4266	0.5088	-0.0537
3833	446	-1	-0.2628	1.5538	0.9268
3850	-1	-1	-0.1082	0.5088	-0.1626

	nueroticism	openness_to_experience
7	0.14590	1.2470
44	0.29738	0.1864
138	1.16210	0.4805
187	0.00353	0.1864
477	1.85433	-0.1295
614	-0.48788	0.8183
698	-0.99500	-0.2859
788	0.17988	-1.0774
874	-1.23827	-0.1295
987	1.50180	-1.2354
1829	0.06223	0.1864
1134	-0.52538	0.9763
1264	-0.17270	0.5824
1345	0.01920	-0.6692
1419	-1.75560	-0.4776
1439	0.41480	-1.15513
1510	-0.26090	-1.0774
1585	0.77980	0.8973
1767	-0.86820	-0.8668
2151	1.47240	0.8183
2152	-2.17840	0.8973
2229	-0.61470	-0.4776

OMATICS%20OCTOBER/project1_innomatics%20NEW.ipynb

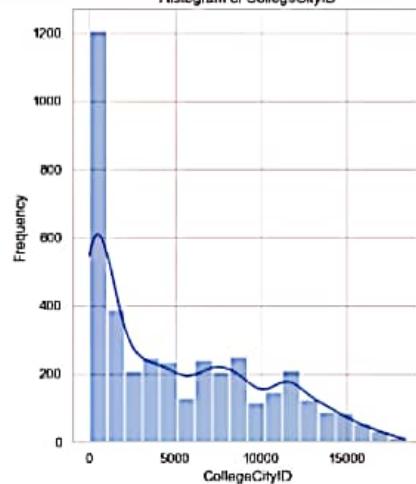


jupyter project1_innomatics NEW Last Checkpoint 30 minutes ago

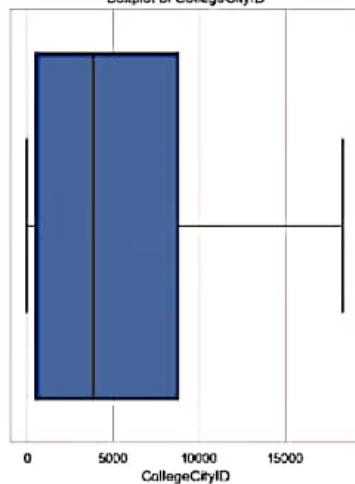
File Edit View Run Kernel Settings Help

+ X D ▶ C ▶ Code

Histogram of CollegeCityID



Boxplot of CollegeCityID

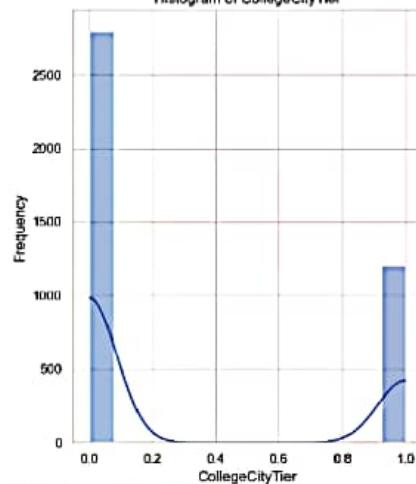
Outliers in CollegeCityID
JupyterLab Python 3 (ipykernel)

Outliers in CollegeCityID:

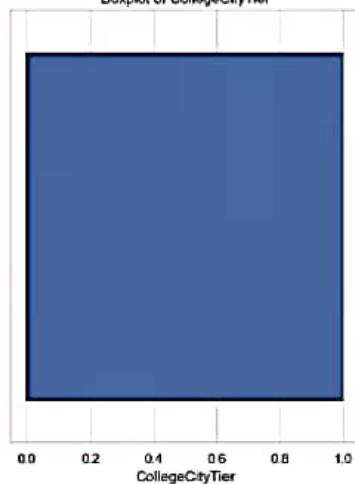
Empty DataFrame
Columns: [Unnamed: 0, ID, Salary, DOJ, DOL, Designation, JobCity, Gender, DOB, 10percentage, 10board, 12graduation, 12percentage, 12board, CollegeID, CollegeTier, Degree, Specialization, collegeGPA, CollegeCityID, CollegeCityTier, CollegeState, GraduationYear, English, Logical, Quant, Domain, ComputerProgramming, ElectronicsAndSemicon, ComputerScience, MechanicalEngg, ElectricalEngg, TelecomEngg, CivilEngg, conscientiousness, agreeableness, extraversion, neuroticism, openess_to_experience]
Index: []

[8 rows x 39 columns]

Histogram of CollegeCityTier



Boxplot of CollegeCityTier



Outliers in CollegeCityTier

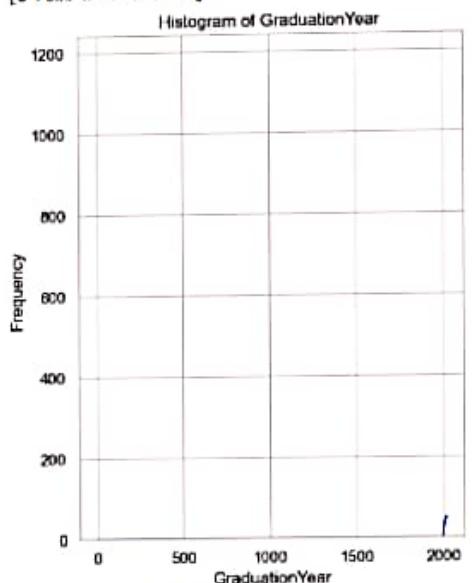


Outliers in CollegeCityTier:

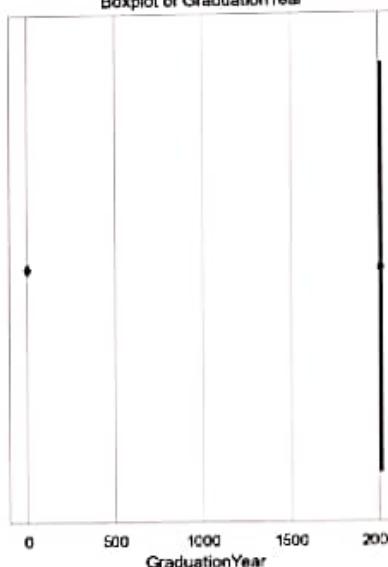


Scanned with OKEN Scanner

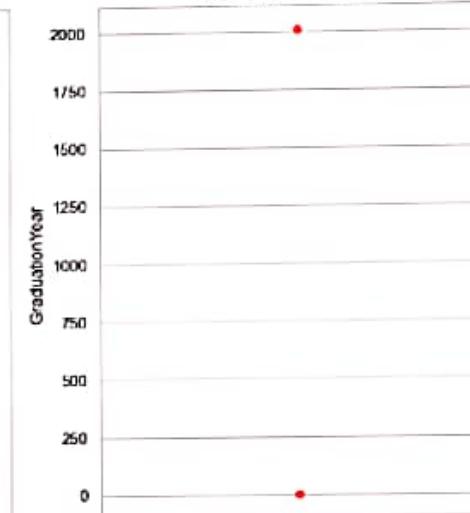
[8 rows x 39 columns]



Boxplot of GraduationYear



Outliers in GraduationYear



Outliers in GraduationYear:

```
      Unnamed: 0      ID   Salary      DOJ      DOL \
2664    train  794381  325000 2014-07-01  2015-07-01
3193    train  480729  120000 2007-06-01  2009-06-01
```

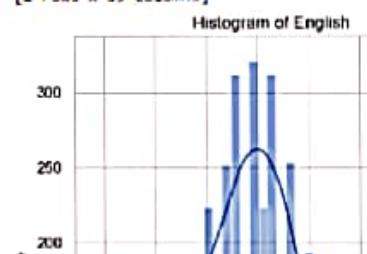
```
      Designation   JobCity Gender      DOB 10percentage ...
2664  systems engineer  Chandigarh     m 1991-12-17      89.8 ...
3193  software engineer  Hyderabad     m 1986-06-18      83.5 ...
```

```
      ComputerScience MechanicalEngg ElectricalEngg TelecomEngg CivilEngg \
2664           -1          446         -1        -1        -1
3193           -1          -1         -1        -1        -1
```

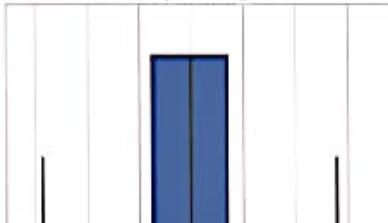
```
      conscientiousness agreeableness extraversion neuroticism \
2664           1.1336      0.3789     -0.2974     -0.3612
3193          -1.4992     -0.7863     -2.8131      0.4148
```

```
      openness_to_experience
2664           -1.2448
3193           -0.6835
```

[2 rows x 39 columns]



Boxplot of English



Outliers in English

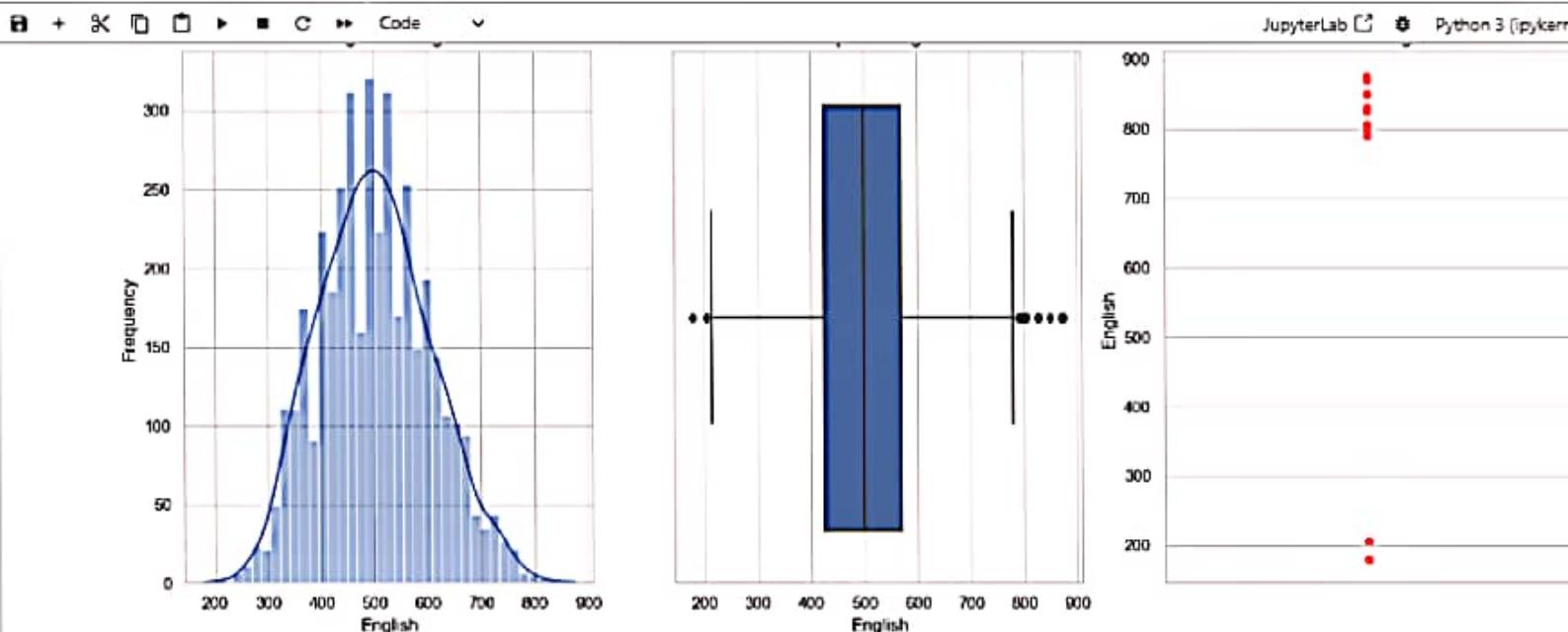


Project 1 - Innomatics

jupyter project1_innomatics NEW Last Checkpoint: 30 minutes ago

File Edit View Run Kernel Settings Help

Not Tr...



Outliers in English:

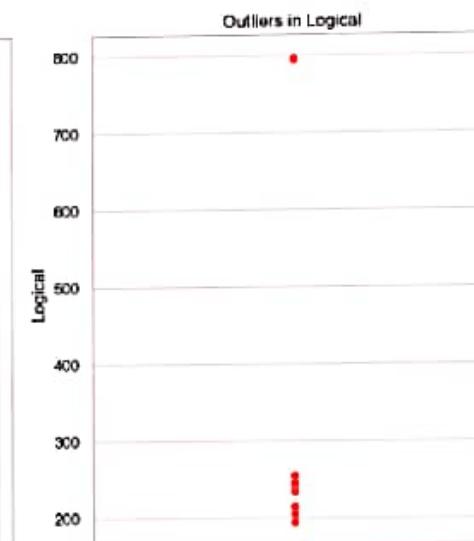
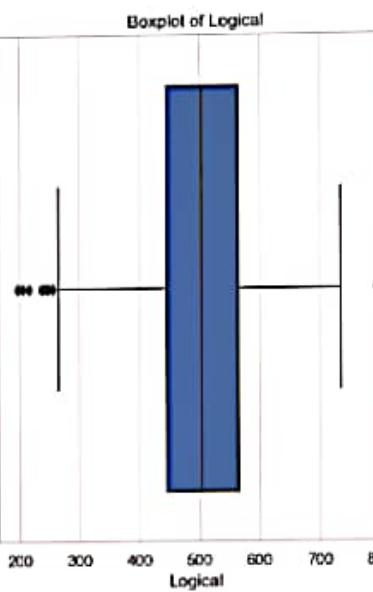
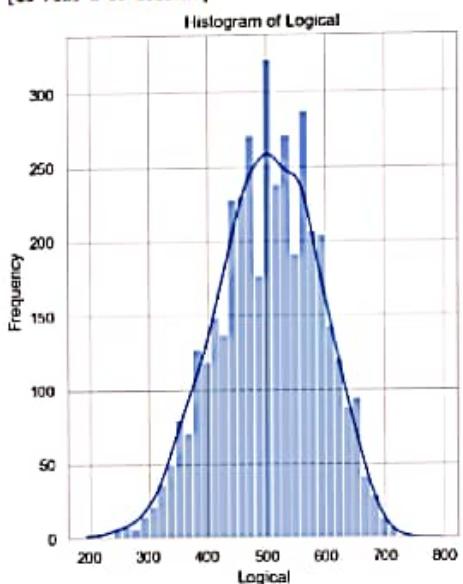
```
Unnamed: 0 ID Salary DOJ DOL \
275 train 1823622 365000 2014-08-01 present
444 train 283164 715000 2012-01-01 present
668 train 609843 440000 2013-10-01 present
839 train 693533 315000 2013-08-01 2015-02-01 00:00:00
847 train 608938 600000 2013-07-01 2014-06-01 00:00:00
935 train 513149 150000 2012-09-01 2013-08-01 00:00:00
1183 train 1883837 310000 2014-09-01 present
1217 train 669692 600000 2014-10-01 2015-02-01 00:00:00
1450 train 1284437 200000 2015-01-01 2015-07-01 00:00:00
1519 train 984550 320000 2014-10-01 present
2077 train 922684 305000 2014-09-01 2015-04-01 00:00:00
2122 train 605290 180000 2012-04-01 2013-02-01 00:00:00
2385 train 683182 325000 2014-02-01 present
2596 train 1088319 610000 2014-07-01 2015-04-01 00:00:00
3844 train 837923 360000 2013-12-01 2015-03-01 00:00:00

Designation JobCity Gender DOB \
275 software engineer -1 m 1993-03-01
444 senior engineer -1 m 1989-06-15
668 software engineer -1 f 1991-09-06
839 hardware engineer Mumbai m 1989-10-02
847 manager Gurgaon m 1991-05-25
935 asp.net developer -1 m 1998-10-24
1183 quality assurance test engineer -1 m 1992-05-19
1217 assistant manager kolkata m 1992-05-04
1450 web designer and seo Bikaner f 1992-11-16
1519 software engineer New Delhi m 1992-09-05
2077 programmer analyst Coimbatore f 1991-05-03
2122 project engineer Nalagarh m 1998-11-03
2385 systems engineer Mysore f 1991-03-14
2596 associate software engineer Hyderabad m 1999-10-04
3844 junior engineer -1 m 1993-07-07

10percentage ... ComputerScience MechanicalEngg ElectricalEngg \
275 88.6 ... 438 -1 -1
```

	nueroticism	openness_to_experience
275	0.8192	0.2889
444	0.8848	0.5024
668	1.9287	0.6721
839	-0.9958	0.4885
847	-0.2344	-0.4776
935	-0.2689	-0.2875
1183	0.3995	-1.2440
1217	0.0192	-2.5853
1458	-1.3753	0.4885
1519	-0.4879	0.6721
2077	-2.1160	0.2889
2122	-0.4879	1.2478
2385	-1.2486	0.0973
2596	-1.2486	1.0554
3844	-1.1218	-0.6692

[15 rows x 39 columns]



Outliers in Logical:

	Unnamed: 0	ID	Salary	DOJ	DOL
101	train	43693	285000	2011-12-01	present
133	train	434552	300000	2014-06-01	2015-04-01 00:00:00
207	train	457244	360000	2013-04-01	present
345	train	68315	1180000	2011-07-01	2012-05-01 00:00:00
628	train	47805	505000	2010-12-01	2014-06-01 00:00:00
1014	train	649530	350000	2014-02-01	present
1160	train	65483	325000	2012-04-01	present
1439	train	299447	360000	2011-08-01	present
2141	train	348845	1000000	2013-04-01	2013-09-01 00:00:00
2265	train	126975	300000	2010-09-01	2011-10-01 00:00:00
2796	train	1087183	1000000	2015-02-01	2015-03-01 00:00:00
2838	train	133458	1500000	2010-10-01	2011-12-01 00:00:00
2891	train	205633	500000	2011-09-01	present
3119	train	961011	95000	2014-04-01	2015-04-01 00:00:00
3159	train	54007	1300000	2010-12-01	2011-12-01 00:00:00
3704	train	21023	1000000	2010-03-01	present

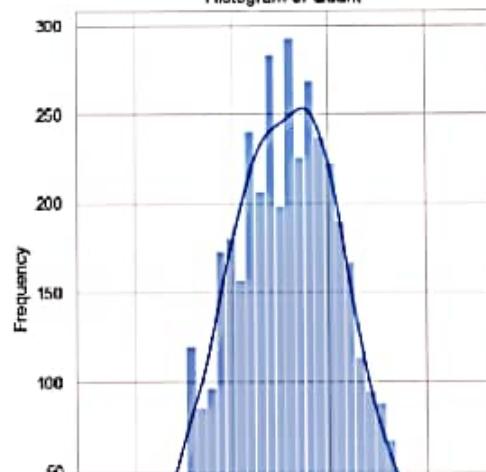
```
CivilEng conscientiousness agreeableness extraversion neuroticism \
101      -1     -0.1082    0.8328   -0.1988    0.29730
133      -1     -0.8810    -0.2793    0.9623    1.11990
207      -1     -2.1175    -2.9314   -2.2388   -0.87780
345      -1     0.2099    0.5008   -1.0697   -0.52530
628      -1     -2.2728    -0.7993   -1.3599    0.53230
1014     -1     -0.4453    1.3779    0.9322   -0.99500
1160     -1     0.5108    -1.0593   -1.2148    0.41480
1439     268    -2.7357    -2.3873   -0.7794    0.41488
2141     -1     -0.8810    0.1888   -0.7794    0.00353
2265     -1     -1.3447    -0.5913   -1.5051   -0.29820
2796     -1     0.9980    -0.9511    1.0859    0.52620
2830     -1     -2.2720    -2.7754   -1.6502    0.41480
2891     -1     -1.3447   -1.3711   -0.4891    0.17980
3119     -1     -0.7335    -2.9511   -0.9122   -0.36120
3159     -1     -0.8810    0.1888   -0.1988   -1.23830
3784     -1     -0.5719    -1.2543    0.1640    1.11983
3876     -1     -2.8983    -0.1232    0.3817   -0.87780
3953     -1     -1.0355    -2.0733   -1.2511    0.17983
```

openness_to_experience

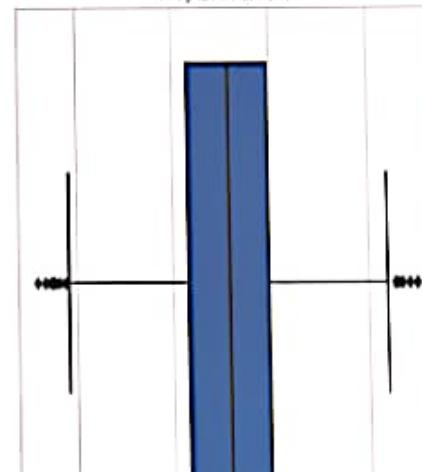
```
101      -0.7615
133      0.1864
207      -3.9685
345      0.5683
628      -2.1833
1014     0.8637
1160     -0.6035
1439     -1.5513
2141     -0.6035
2265     -0.7615
2796     -0.8943
2830     -2.3412
2891     1.2923
3119     -5.8428
3159     0.3444
3784     -1.3934
3876     -0.6035
3953     -1.5513
```

[18 rows x 39 columns]

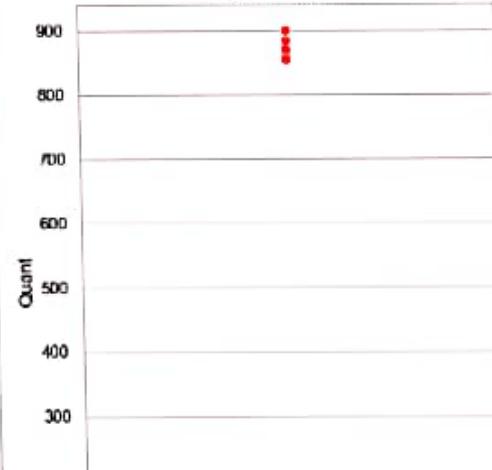
Histogram of Quant

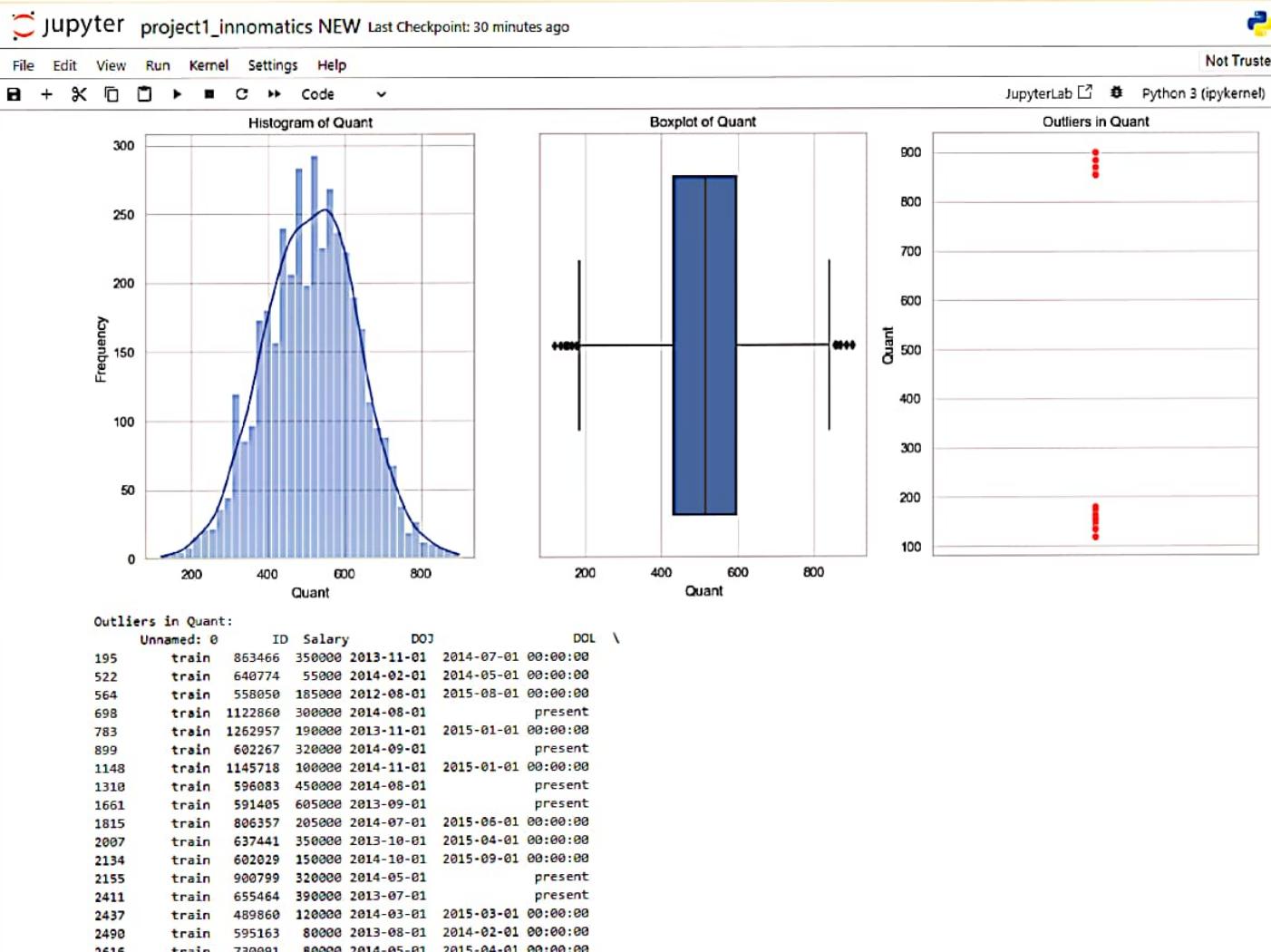


Boxplot of Quant



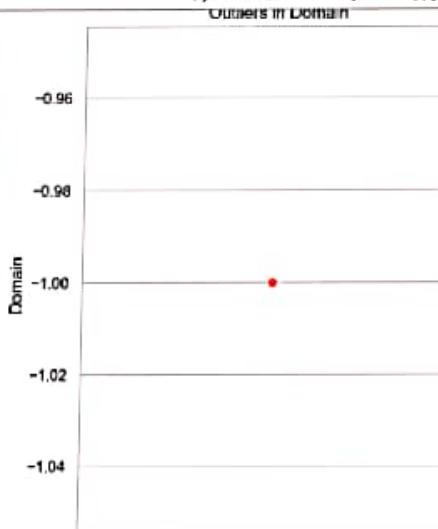
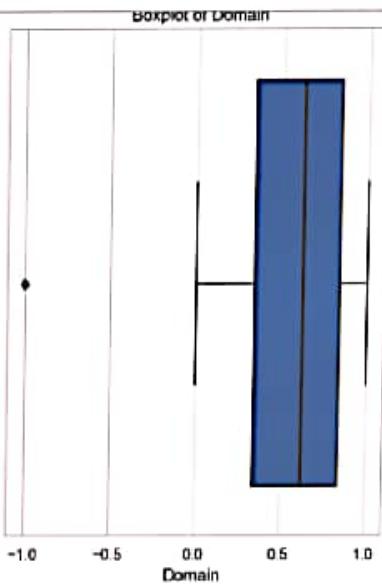
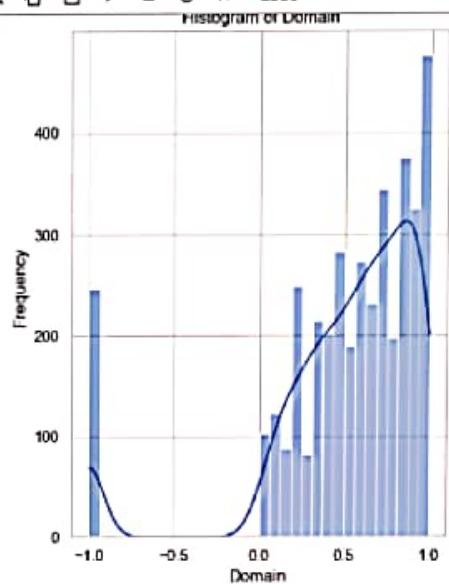
Outliers in Quant





File Edit View Run Kernel Settings Help

+ X C Code

JupyterLab Python 3 (ipykernel)
Outliers in Domain

Outliers in Domain:

```
      Unnamed: 0   ID  Salary      DOB      DOL \
5       train  1027655  300000  2014-06-01    present
14      train  1279958  300000  2013-07-01    present
16      train  1087625  325000  2014-09-01    present
22      train  995191  190000  2013-10-01 2015-04-01 00:00:00
35      train  942679  350000  2014-03-01    present
...
3942     train  1067815  280000  2013-07-01    present
3958     train  1039828  310000  2014-06-01    present
3960     train  1037712  240000  2013-08-01 2014-01-01 00:00:00
3969     train  1233826  330000  2015-06-01    present
3978     train  1274658  320000  2013-08-01 2014-01-01 00:00:00
```

```
      Designation  JobCity Gender      DOB 10percentage \
5        system engineer  Hyderabad  m 1992-07-02    89.92
14      java software engineer  Bangalore  m 1992-07-02    81.20
16  product development engineer  Noida  m 1992-06-10    85.80
22        developer  Hyderabad  m 1992-01-23    92.50
35  online marketing manager  mohali  f 1998-09-15    85.00
...
3942      software engineer     -1  m 1998-03-17    78.00
3958      software engineer  Hyderabad  m 1992-09-09    79.90
3960        web developer  Kolkata  m 1992-07-10    77.00
3969      technical engineer  pune  m 1993-01-24    76.00
3978      software engineer  Hyderabad  f 1991-04-24    89.70
```

```
... ComputerScience  MechanicalEngg  ElectricalEngg  TelecomEngg \
5 ...           487         -1          -1          -1
14 ...           376         -1          -1          -1
16 ...           -1         -1          -1          -1
22 ...           -1         -1          -1          -1
35 ...           -1         -1          -1          -1
...
3942 ...           346         -1          -1          -1
3958 ...           438         -1          -1          -1
3960 ...           -1         -1          -1          -1
```



File Edit View Run Kernel Settings Help

Not Trusted

JupyterLab ⚡ Python 3 (ipykernel) C

```

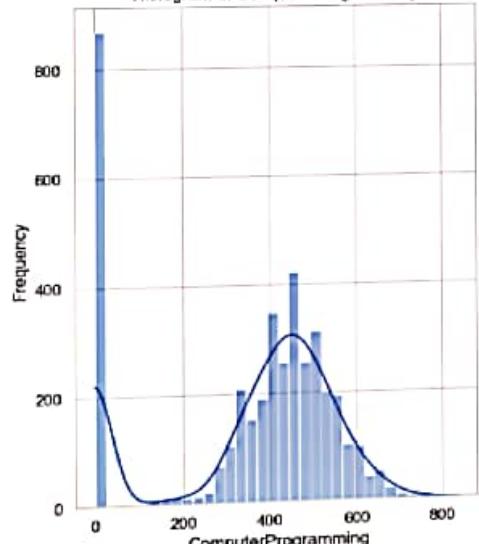
3978 ... -1 -1 -1 -1 -1
CivilEngg conscientiousness agreeableness extraversion neuroticism \
5 -1 -0.3027 -0.6201 -2.2954 -0.7415
14 -1 0.7027 0.7119 0.0100 -0.9950
16 -1 1.1336 1.0449 0.7785 -0.6147
22 -1 -0.4463 -0.6201 -0.6848 0.7798
35 -1 -0.4854 -0.5213 2.1129 -0.7350
... ...
3942 -1 1.7881 -0.9531 0.1637 -1.6289
3958 -1 -0.1590 0.5454 0.4711 -0.1076
3960 -1 0.5591 0.7119 0.4711 -1.6289
3969 -1 0.5591 0.7119 0.0100 -0.2344
3978 -1 0.2718 1.2114 0.4711 -0.1076

openness_to_experience
5 -0.8688
14 0.6721
16 -0.8688
22 -2.7769
35 1.8224
...
3942 -0.8688
3958 0.8637
3960 0.6721
3969 0.8637
3978 0.8637

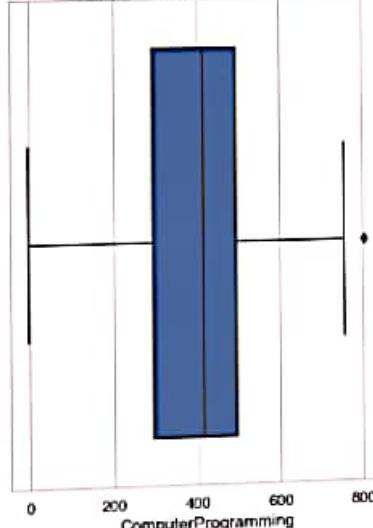
```

[246 rows x 39 columns]

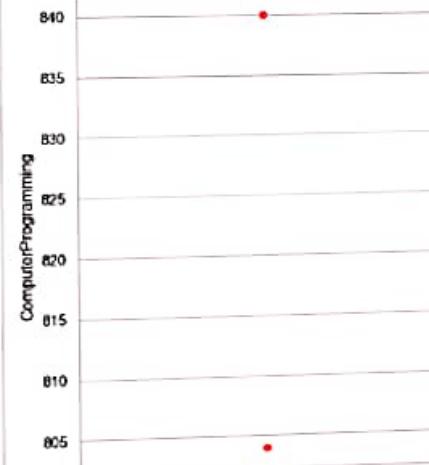
Histogram of ComputerProgramming



Boxplot of ComputerProgramming



Outliers in ComputerProgramming



Outliers in ComputerProgramming:

```

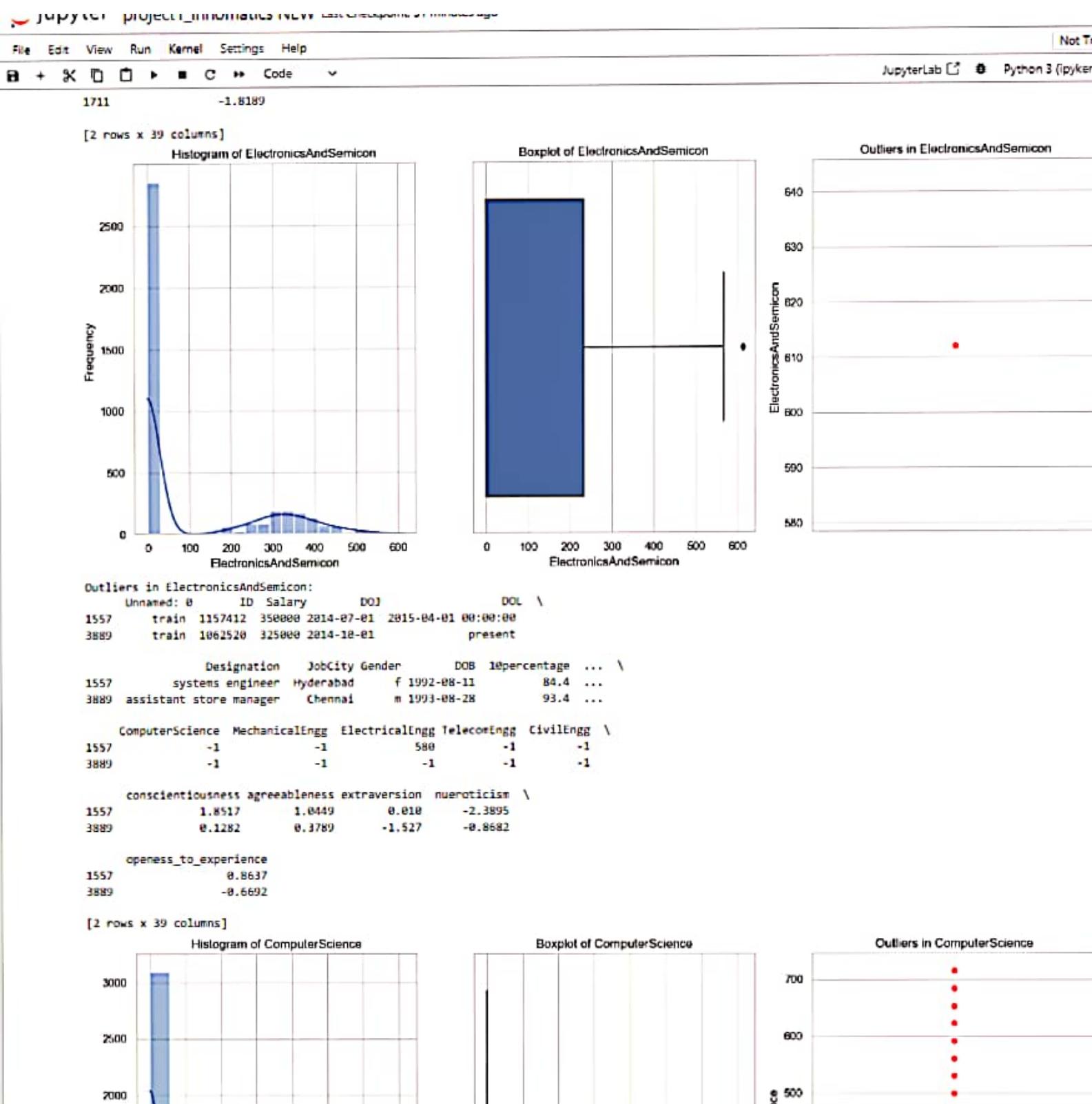
Unnamed: 0 ID Salary DOB \
64 train 31330 600000 2012-07-01 2012-07-01 00:00:00
64 train 712578 625000 2014-02-01 present
1711 train 712578 625000 2014-02-01

Designation JobCity Gender DOB 10percentage ...
64 java software engineer Gurgaon m 1986-05-29 78.0 ...
1711 data analyst -1 m 1998-06-24 88.0 ...

```



Scanned with OKEN Scanner



jupyter project1_innomatics NEW Last Checkpoint 31 minutes ago

File Edit View Run Kernel Settings Help Not Trust

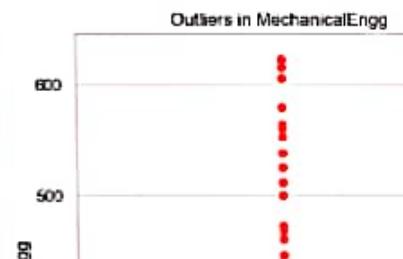
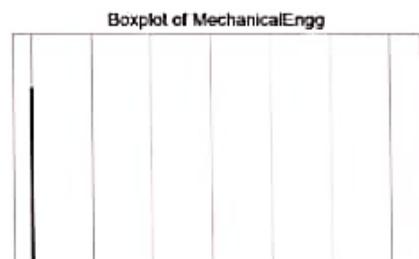
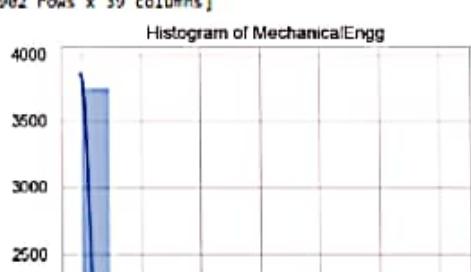
JupyterLab Python 3 (ipykernel)

```
5           system engineer      Hyderabad   m 1992-07-02
6           java software engineer    Bangalore   m 1993-02-01
14          java software engineer    Bangalore   m 1992-07-02
17          associate software developer    Gurgaon    m 1987-09-28
24          programmer analyst      Hyderabad   m 1993-06-28
...
...
3972         help desk analyst      Noida       m 1992-11-30
3976         graduate engineer trainee    Navi Mumbai f 1992-09-13
3981         software engineer        Gurgaon    m 1991-12-17
3989         software engineer        Bangalore   m 1991-11-23
3996         software developer       Asifabadbanglore f 1992-01-20

10percentage ... ComputerScience MechanicalEngg ElectricalEngg \
5           89.92 ...           487          -1          -1
6           86.08 ...           346          -1          -1
14          81.20 ...           376          -1          -1
17          68.80 ...           500          -1          -1
24          88.00 ...           346          -1          -1
...
...
3972         68.28 ...           487          -1          -1
3976         64.28 ...           561          -1          -1
3981         53.48 ...           530          -1          -1
3989         74.88 ...           346          -1          -1
3996         78.72 ...           438          -1          -1

TelecomEngg CivilEngg conscientiousness agreeableness extraversion \
5           -1          -1          -0.3827     -0.6201     -2.2954
6           -1          -1          1.7081      -0.1054     -1.0379
14          -1          -1          0.7827      0.7119      0.0180
17          -1          -1          0.8453      0.8449     -1.0659
24          -1          -1          0.4155      0.8827      0.1357
...
...
3972         -1          -1          -1.1644     0.2124     -0.2974
3976         -1          -1          0.4155      0.0459      0.0180
3981         -1          -1          0.1282     -0.2871     -0.1437
3989         -1          -1          0.1282      0.0459      1.2396
3996         -1          -1          -0.1590     0.0459     -0.4511

nueroticism openness_to_experience
5           -0.7015      -0.8688
6           -2.0092      -1.8872
14          -0.9958      0.6721
17          -1.6289      0.6721
24          -0.9958      -0.6692
...
...
3972         0.0192      -1.0524
3976         0.2727      0.0973
3981         -1.1218      1.4386
3989         1.0333      0.6721
3996         -0.3612     -0.0943
```



jupyter project1_innomatics NEW Last Checkpoint: 31 minutes ago



File Edit View Run Kernel Settings Help

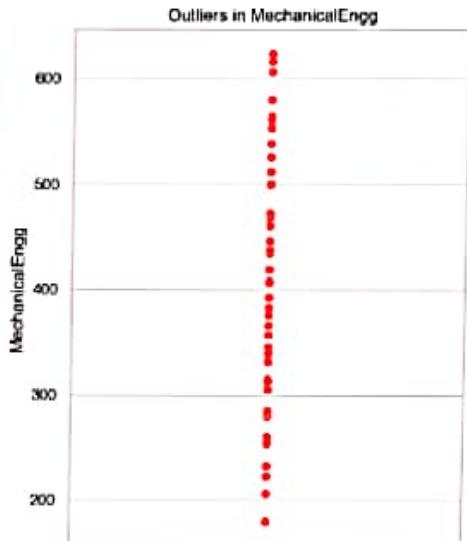
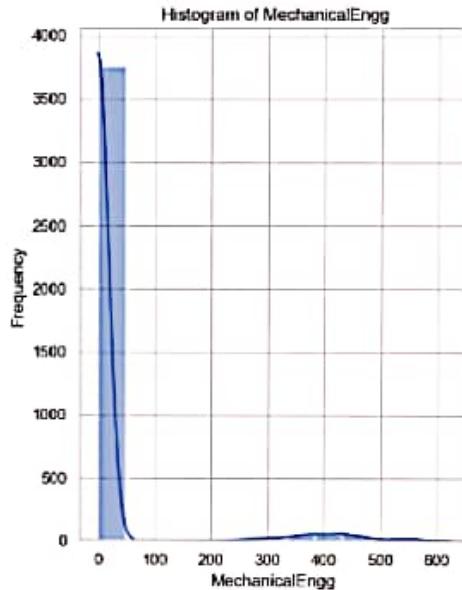
Not Trusted

+

Boxplot of MechanicalEngg

JupyterLab Python 3 (ipykernel)

[982 rows x 39 columns]



Outliers in MechanicalEngg:

```
Unnamed: 0    ID  Salary      DOB      DOL \
7     train  912934  400000  2014-07-01  2015-07-01 00:00:00
26    train   65006  345000  2011-01-01  2013-10-01 00:00:00
40    train  245887  120000  2011-06-01  2012-03-01 00:00:00
45    train  1230582  118000  2014-09-01  2015-01-01 00:00:00
59    train  536053  120000  2009-09-01  2013-04-01 00:00:00
...
3989   train  628271  210000  2013-08-01        present
3916   train  530146  360000  2013-06-01  2014-12-01 00:00:00
3929   train  336917  180000  2012-06-01  2014-04-01 00:00:00
3951   train  885459  145000  2014-05-01        present
3980   train  197796  150000  2011-07-01  2012-07-01 00:00:00
```

```
Designation      JobCity Gender      DOB 10percentage \
7   mechanical engineer  Bangalore   m 1992-05-27    92.00
26 senior software engineer  Bangalore   m 1988-02-05    98.88
40       site manager      Jhansi   m 1989-05-04    66.60
45   production engineer      Chennai   m 1992-10-12    95.80
59   software engineer  Bangalore   m 1977-10-30    72.00
...
3989   production engineer  Navi Mumbai   m 1992-10-25    67.00
3916   management trainee      Hissar   m 1991-07-14    79.80
3929       cad designer      Coimbatore   m 1998-08-24    70.27
3951   mechanical engineer      Karad   m 1992-08-27    93.07
3980         get      haryana   m 1986-08-05    84.00
```

```
... ComputerScience  MechanicalEngg  ElectricalEngg  TelecomEngg \
7 ...
26 ...
40 ...
45 ...
59 ...
...
3989 ...
...
```



Scanned with OKEN Scanner

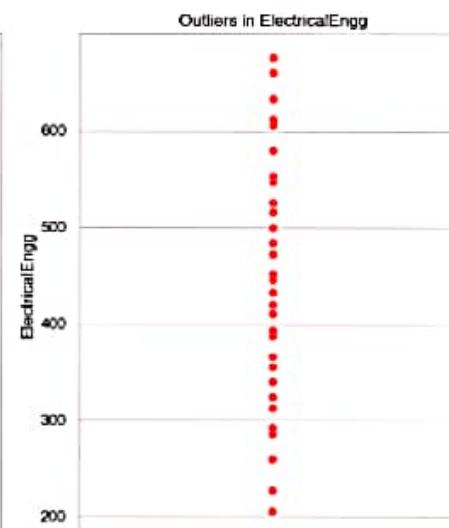
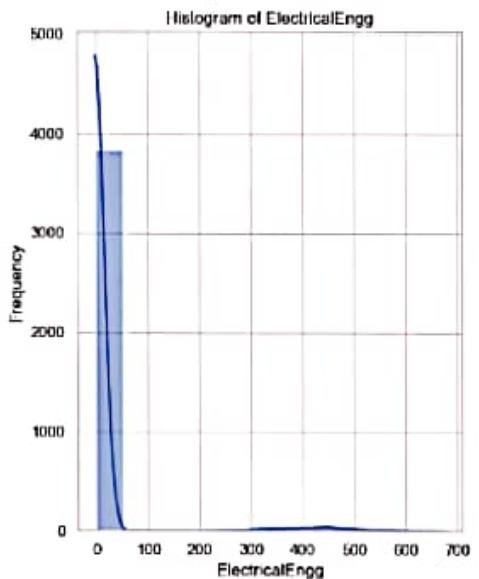
```

3980 ... -1 473 -1 -1
    CivilEngg conscientiousness agreeableness extraversion neuroticism \
7 -1 -0.0154 1.2114 0.8188 0.14598
26 320 0.8192 0.1888 -1.7954 1.82498
48 -1 -0.4173 0.3448 0.0914 -0.40788
45 -1 -0.5899 -1.1196 0.4711 -0.74158
59 480 -0.3014 -0.4353 -0.9245 -0.34988
... ...
3989 -1 0.7827 -0.0873 -0.0319 0.65388
3916 -1 0.3941 0.1888 -0.6343 0.89178
3929 -1 -0.2628 0.5088 -1.0334 0.88488
3951 -1 -0.1027 -1.4526 -0.2974 1.28698
3980 -1 -1.4992 0.5088 -0.9245 -0.64288

openess_to_experience
7 1.2478
26 1.1343
48 0.6603
45 -1.4356
59 -0.4455
... ...
3989 0.1678
3916 -1.3934
3929 -0.4455
3951 -1.4356
3980 -0.6035

```

[235 rows x 39 columns]



Outliers in ElectricalEngg:

```

Unnamed: 0 ID Salary DOJ      DOL \
53   train  1882717 3500000 2014-09-01  present
70   train  801755  3200000 2014-03-01  present
124  train  614524  3150000 2014-11-01  present
125  train  796063  3000000 2014-11-01  2015-03-01 00:00:00
180  train  639785  1800000 2013-09-01  2014-04-01 00:00:00

```

jupyter project1_innomatics NEW Last Checkpoint 31 minutes ago

File Edit View Run Kernel Settings Help



Not Trusted

JupyterLab Python 3 (ipykernel)

```
Designation JobCity Gender DOB 10percentage \
53  assistant system engineer Mumbai m 1991-08-06 89.25
78  system engineer Chennai m 1992-12-12 89.00
124  programmer analyst -1 f 1991-02-06 82.30
125  programmer analyst Kolkata m 1992-03-19 98.00
188  design engineer New Delhi m 1991-11-16 65.80
...
... ...
... ...
3832  engineer Haldia m 1991-10-28 92.83
3835  project engineer Kolkata m 1992-12-12 76.00
3837  application developer Bangalore f 1992-01-26 72.30
3839  ase bhubaneswar f 1992-03-11 98.80
3844  electrical engineer Bangalore m 1992-07-28 90.56

... ComputerScience MechanicalEngg ElectricalEngg TelecomEngg \
53 ... -1 -1 484 -1
78 ... -1 -1 606 -1
124 ... -1 -1 393 -1
125 ... -1 -1 500 -1
188 ... -1 -1 553 -1
...
... ...
... ...
3832 ... -1 -1 660 -1
3835 ... -1 -1 356 -1
3837 ... -1 -1 366 -1
3839 ... -1 -1 686 -1
3844 ... 284 -1 588 -1

CivilEngg conscientiousness agreeableness extraversion neuroticism \
53 -1 -0.1500 -0.1206 -0.6048 -1.6289
78 -1 -1.4517 -1.4526 -1.2196 1.7948
124 -1 -0.1590 -1.1195 1.2396 1.6672
125 -1 0.1282 -0.2871 1.3933 0.6538
188 -1 0.5591 -0.1206 -2.1418 0.7798
...
... ...
... ...
3832 -1 0.1282 0.8459 -0.1437 -0.2344
3835 -1 0.5591 0.3789 -0.4511 0.7798
3837 -1 -0.4463 -1.2861 -2.4491 -1.3753
3839 -1 1.5544 0.5454 -0.2974 -1.3753
3844 -1 -0.1500 0.8459 -0.1437 -0.7415

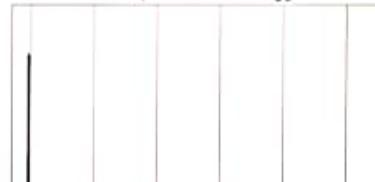
openess_to_experience
53 -1.4356
78 -2.3937
124 0.6721
125 -0.6692
188 -1.0524
...
... ...
3832 0.8973
3835 1.0554
3837 -0.2859
3839 1.0554
3844 0.4805
```

[161 rows x 39 columns]

Histogram of TelecomEngg



Boxplot of TelecomEngg

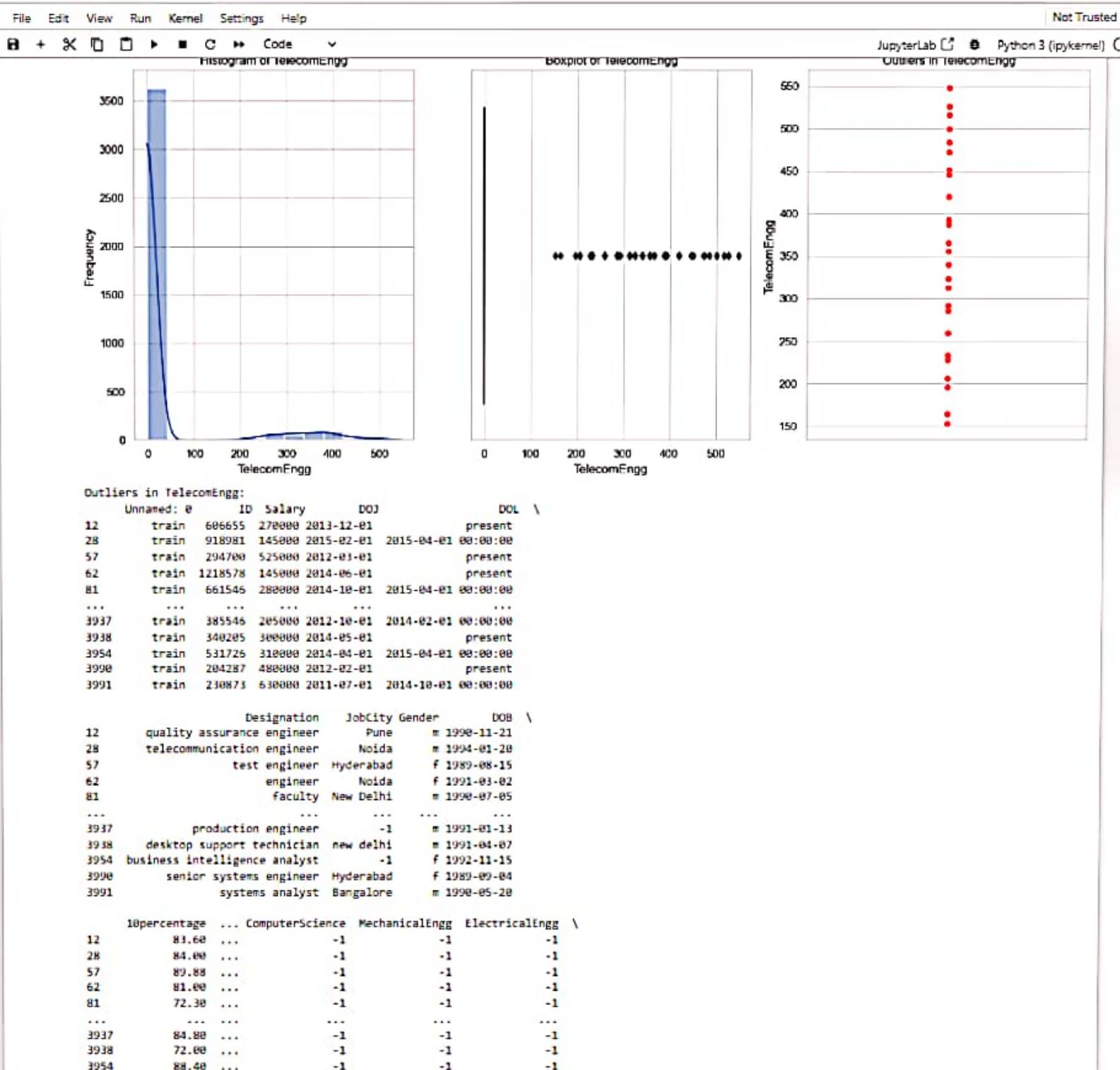


Outliers in TelecomEngg



Project 1: Exploratory Data Analysis

jupyter project1_innomatics NEW Last Checkpoint 31 minutes ago



Outliers in TelecomEngg:

```
    Unnamed: 0   ID  Salary      DOJ      DOL \
12      train  606655 270000 2013-12-01  present
28      train  918981 145000 2015-02-01 2015-04-01 00:00:00
57      train  294700 525000 2012-03-01  present
62      train  1218578 145000 2014-06-01  present
81      train  661545 280000 2014-10-01 2015-04-01 00:00:00
...     ...   ...
3937     train  385546 285000 2012-10-01 2014-02-01 00:00:00
3938     train  340205 300000 2014-05-01  present
3954     train  531726 310000 2014-04-01 2015-04-01 00:00:00
3990     train  204287 480000 2012-02-01  present
3991     train  210873 630000 2011-07-01 2014-10-01 00:00:00
```

```
        Designation      JobCity Gender      DOB \
12  quality assurance engineer      Pune   m 1990-11-21
28  telecommunication engineer      Noida   m 1994-01-28
57          test engineer      Hyderabad   f 1989-08-15
62            engineer      Noida   f 1991-03-02
81            faculty  New Delhi   m 1998-07-05
...     ...   ...
3937      production engineer       -1   m 1991-01-13
3938  desktop support technician  new delhi   m 1991-04-07
3954 business intelligence analyst       -1   f 1992-11-15
3990      senior systems engineer      Hyderabad   f 1989-09-04
3991      systems analyst      Bangalore   m 1998-05-28
```

```
10percentage ... ComputerScience MechanicalEngg ElectricalEngg \
12      83.60   ...      -1      -1      -1
28      84.00   ...      -1      -1      -1
57      89.88   ...      -1      -1      -1
62      81.00   ...      -1      -1      -1
81      72.38   ...      -1      -1      -1
...     ...   ...
3937     84.88   ...      -1      -1      -1
3938     72.00   ...      -1      -1      -1
3954     88.40   ...      -1      -1      -1
```



File Edit View Run Kernel Settings Help

Not Trusted

JupyterLab ○ Python 3 (ipykernel) ○

```

3991    80.00 ... -1    -1    -1

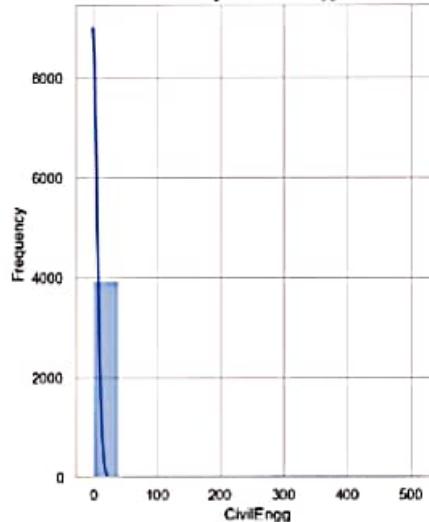
    TelecomEngg CivilEngg conscientiousness agreeableness extraversion \
12      285     -1     0.1282    0.5454   -0.7585
28      313     -1     0.1282    0.7119   1.3933
57      420     -1     0.3555   -0.2793   -0.6343
62      268     -1    -1.1644   -0.9511    0.6198
81      393     -1     1.4288    0.5454   -0.2974
...
...    ...
3937    233     -1     0.1623    0.0328   -0.3440
3938    348     -1    -0.5719    0.5808   -0.0537
3954    393     -1     0.8192    0.6568   0.9623
3990    233     -1     0.6646    0.3448   0.3817
3991    393     -1    -1.3447    0.5808   -1.6502

nueroticism openness_to_experience
12     -0.99588    0.8537
28     -0.36128   -0.0943
57     -0.29828   -0.2875
62     -0.74158   -0.8608
81     1.03338   -0.6692
...
...
3937    -1.05418    0.5824
3938    1.11983    1.2528
3954    -0.29828    0.8679
3990    -1.34780    0.8183
3991    -0.05520    0.8284

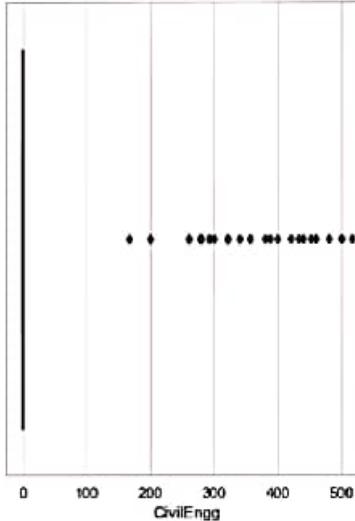
```

[374 rows x 39 columns]

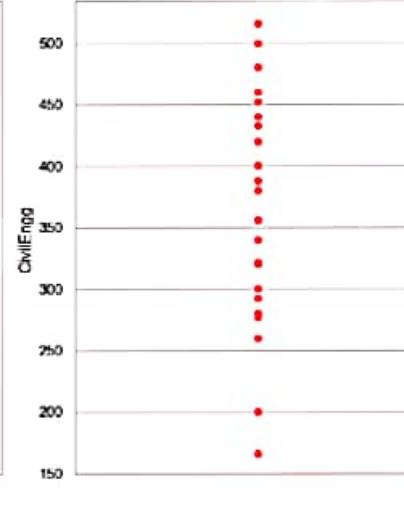
Histogram of CivilEngg



Boxplot of CivilEngg



Outliers in CivilEngg



Outliers in CivilEngg:

```

        Unnamed: 0   ID  Salary      DOJ      DOL \
26      train    65896  345000 2011-01-01  2013-10-01 00:00:00
59      train    536853 1200000 2009-09-01  2013-04-01 00:00:00
172     train   1889283 1100000 2014-06-01  2015-02-01 00:00:00
211     train    46302  350000 2010-10-01  2013-05-01 00:00:00
222     train   226736 1000000 2013-03-01  2013-06-01 00:00:00
321     train    24114  200000 2010-04-01  2011-01-01 00:00:00

```



Scanned with OKEN Scanner

jupyter project1_innomatics NEW Last Checkpoint: 31 minutes ago



File Edit View Run Kernel Settings Help

Not Trusted

Back Forward Stop Refresh Code

JupyterLab Python 3 (ipykernel)

	Designation	JobCity	Gender	DOB	\
26	senior software engineer	Bangalore	m	1988-02-05	
59	software engineer	Bangalore	m	1977-10-30	
172	assistant professor	Una	f	1993-04-15	
211	maintenance engineer	Nagpur	m	1988-07-31	
222	programmer	Coimbatore	m	1989-05-13	
331	assistant manager	pune	f	1990-02-16	
483	planning engineer	Noida	m	1991-10-06	
453	software engineer	pune	m	1992-10-29	
492	business analyst	Noida	m	1992-09-11	
493	data analyst	Gurgaon	m	1991-08-24	
505	application engineer	New Delhi	m	1988-07-13	
605	software test engineer	Bangalore	m	1990-12-07	
687	engineer	New Delhi	m	1992-07-01	
896	project engineer	Hyderabad	f	1989-08-02	
1108	assistant engineer	Dehradun	m	1993-09-16	
1217	assistant manager	kolkata	m	1992-05-04	
1253	database administrator	Hyderabad	m	1990-03-07	
1282	software engineer	Bangalore	m	1989-12-15	
1439	assistant professor	AM	m	1988-11-11	
1581	technical engineer	New Delhi	m	1993-07-10	
1678	management trainee	Pune	f	1992-05-10	
1858	project engineer	-1	m	1992-08-05	
1879	junior research fellow	Chandigarh	m	1992-03-18	
1880	technical support specialist	-1	m	1989-09-14	
2242	assistant manager	PATNA	m	1993-01-20	
2363	product development engineer	Tirupati	m	1988-11-08	
2419	design engineer	Pune	m	1992-05-10	
2469	civil engineer	Joshimath	m	1991-09-28	
2571	business analyst	Bangalore	f	1992-09-27	
2591	lecturer	Chennai	m	1991-12-27	
2698	civil engineer	Ahmedabad	m	1992-10-10	
3015	civil engineer	patna	m	1985-11-17	
3332	design engineer	Chennai	f	1993-02-14	
3513	assistant engineer	Mumbai	m	1992-04-02	
3675	software engineer	Chennai	f	1989-05-23	
3790	embedded software engineer	Bangalore	m	1987-07-18	
3821	engineer	Gurgaon	m	1992-01-01	
3865	design engineer	ranchi	m	1992-10-29	
3868	project engineer	jaipur	m	1991-09-14	
3985	system administrator	-1	m	1991-08-28	
3961	planning engineer	Rajpura	m	1987-12-27	
3992	manager	Rajkot	m	1990-06-22	

	10percentage	ComputerScience	MechanicalEngg	ElectricalEngg	\
26	98.88	...	-1	313	-1
59	72.00	...	-1	366	-1
172	70.00	...	-1	-1	-1
211	74.48	...	-1	393	-1
222	77.00	...	-1	268	-1
331	82.50	...	-1	-1	-1
483	93.48	...	-1	-1	-1
453	90.15	...	-1	-1	-1
492	92.00	...	-1	-1	-1
493	87.80	...	-1	-1	-1
505	71.66	...	-1	-1	-1
605	87.11	...	-1	-1	-1
687	92.00	...	-1	-1	-1
896	98.83	...	-1	366	-1
1108	73.00	...	-1	-1	-1
1217	73.00	...	-1	313	-1
1253	79.88	...	-1	366	-1
1282	61.00	...	-1	268	-1



Scanned with OKEN Scanner

Project 1: Innomatics

jupyter project1_innomatics NEW Last Checkpoint: 31 minutes ago

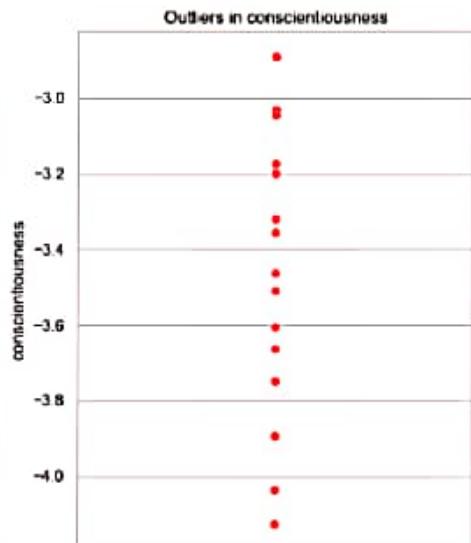
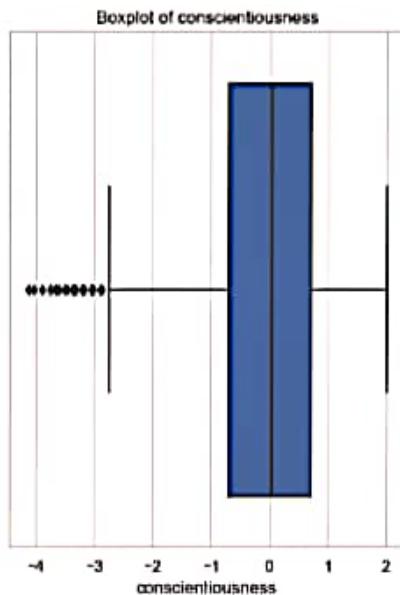
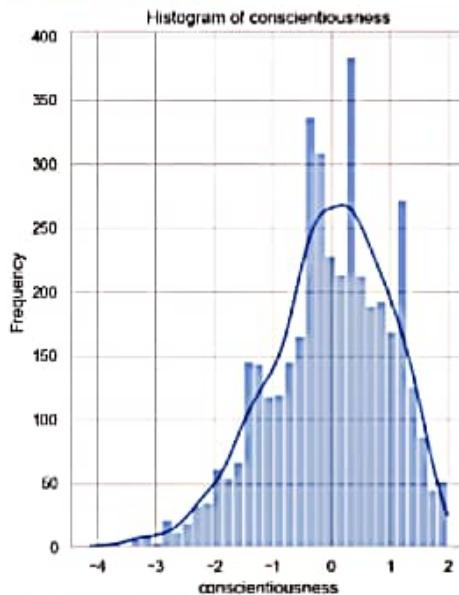
File Edit View Run Kernel Settings Help

Not Trusted

Code

JupyterLab Python 3 (ipykernel)

[42 rows x 39 columns]



Outliers in conscientiousness:

	Unnamed: 0	ID	Salary	DOJ	DOL	\
29	train	472956	220000	2014-11-01	2014-04-01	00:00:00
159	train	755088	340000	2014-05-01	2015-04-01	00:00:00
218	train	38739	310000	2012-09-01	2012-03-01	00:00:00
315	train	232438	360000	2012-05-01		present
335	train	1842343	120000	2014-02-01		present
373	train	215639	370000	2014-03-01		present
382	train	58739	260000	2011-03-01		present
488	train	1175511	280000	2015-02-01	2015-07-01	00:00:00
468	train	1873243	240000	2013-07-01		present
523	train	1061906	120000	2014-03-01	2015-04-01	00:00:00
1211	train	632729	420000	2013-08-01		present
1337	train	355018	285000	2013-02-01		present
1353	train	578644	350000	2013-06-01		present
1684	train	455828	450000	2013-03-01		present
1687	train	326762	450000	2013-03-01	2015-03-01	00:00:00
1972	train	228051	685000	2012-05-01	2014-12-01	00:00:00
2005	train	225059	395000	2013-09-01		present
2046	train	309046	250000	2014-02-01	2015-01-01	00:00:00
2181	train	358268	450000	2011-03-01		present
2182	train	41147	400000	2010-01-01	2011-12-01	00:00:00
2223	train	52923	465000	2011-06-01		present
2224	train	624580	325000	2014-03-01	2015-04-01	00:00:00
2377	train	618785	335000	2014-04-01	2015-04-01	00:00:00
2396	train	995921	240000	2014-01-01	2015-03-01	00:00:00
2569	train	337764	220000	2012-10-01	2015-04-01	00:00:00
3047	train	1283490	95000	2014-11-01		present
3150	train	1104236	145000	2014-07-01	2015-04-01	00:00:00
3372	train	852825	310000	2014-09-01		present
3407	train	242761	600000	2011-06-01		present
3468	train	195483	520000	2011-09-01		present
3473	train	1272992	35000	2014-05-01	2015-05-01	00:00:00
3569	train	519984	630000	2012-11-01		present
3629	train	947114	150000	2014-07-01	2015-01-01	00:00:00

jupyter project1_innomatics NEW Last Checkpoint 32 minutes ago

File Edit View Run Kernel Settings Help

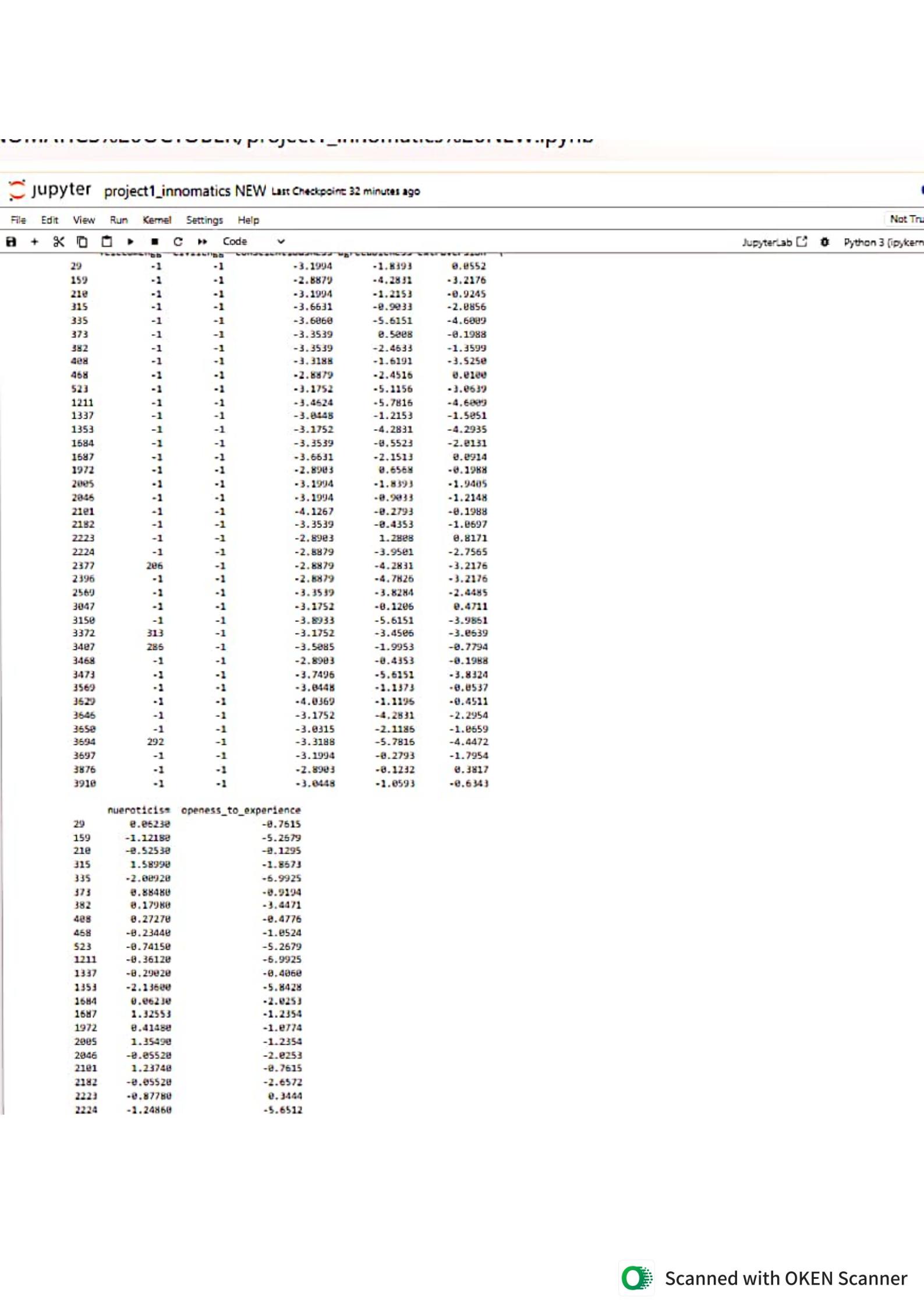
Not Trusted

3918 train 111273 290000 2013-09-01 2014-10-01 00:00:00

JupyterLab Python 3 (ipykernel)

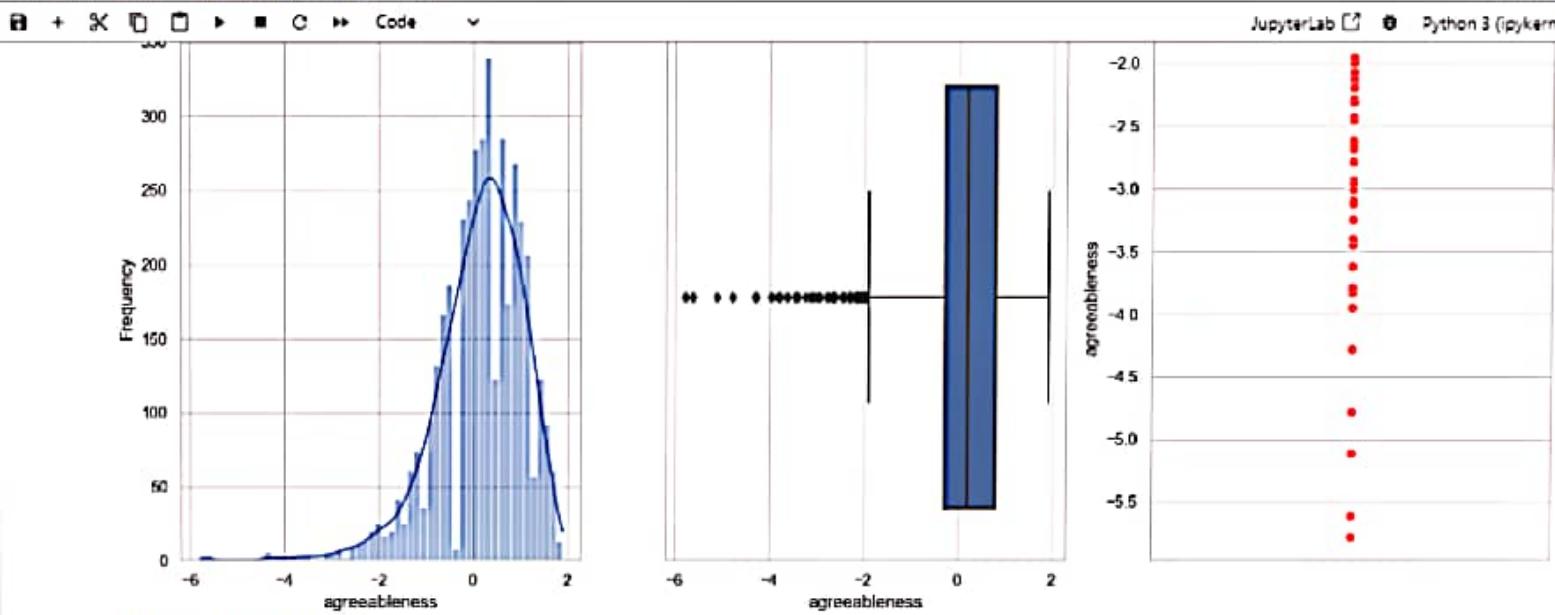
	Designation	JobCity	Gender	DOS
29	java software engineer	Gurgaon	m	1992-06-11
159	asp.net developer	Kolkata	m	1992-03-30
218	network engineer	Lucknow	m	1987-08-15
315	technical engineer	NOIDA	m	1998-07-10
335	electrical engineer	Siliguri	m	1991-03-26
373	assistant manager	Kolkata	m	1989-04-28
382	java software engineer	Bangalore	f	1987-06-26
408	web designer	Gurgaon	m	1993-06-06
468	system engineer	KOLKATA	m	1991-02-24
523	junior software developer	Hyderabad	f	1991-03-13
1211	senior network engineer	Gurgaon	m	1991-08-13
1337	test engineer	Bangalore	m	1991-05-11
1353	quality assurance auditor	Coimbatore	m	1992-09-05
1684	software developer	Hyderabad	m	1989-12-05
1687	system engineer	Pune	m	1991-07-26
1972	software engineer	Bangalore	m	1988-04-04
2005	quality assurance automation engineer	Bangalore	m	1989-11-03
2046	android developer	-1	m	1993-11-21
2101	system engineer	Mumbai	m	1998-05-06
2182	automation engineer	gurgaon	m	1987-09-13
2223	technical lead	Pune	m	1989-08-16
2224	software engineer	Chennai	f	1992-03-22
2377	programmer analyst	Bangalore	m	1991-07-01
2396	software test engineer	Pune	m	1991-01-17
2569	data analyst	Coimbatore	m	1988-07-26
3047	graduate trainee engineer	Faridabad	m	1992-07-10
3150	business analyst	Bangalore	f	1992-10-08
3372	programmer analyst	Pune	m	1992-12-22
3487	senior software engineer	Bangalore	m	1988-04-18
3468	system engineer	Noida	m	1989-11-10
3473	.net developer	hyderabad	m	1993-01-08
3569	software test engineer	Bangalore	m	1990-08-17
3629	java software engineer	Bangalore	m	1992-01-22
3646	software developer	Bangalore	f	1991-06-25
3658	engineer	Pune	m	1991-03-25
3694	rf engineer	Lucknow	m	1992-09-05
3697	software engineer	Hyderabad	m	1989-06-26
3876	software developer	-1	m	1987-06-01
3918	research engineer	Pune	m	1988-11-05

	18percentage	ComputerScience	MechanicalEngg	ElectricalEngg
29	64.00	...	376	-1
159	87.00	...	-1	-1
218	69.00	...	-1	-1
315	88.00	...	-1	-1
335	68.50	...	-1	324
373	81.60	...	446	-1
382	79.89	...	-1	-1
408	68.00	...	253	-1
458	82.00	...	-1	-1
523	63.00	...	-1	-1
1211	88.00	...	-1	-1
1337	65.23	...	-1	-1
1353	85.00	...	473	-1
1584	88.00	...	-1	-1
1587	63.00	...	-1	-1
1972	58.00	...	-1	-1
2005	86.13	...	469	-1
2046	71.00	...	346	-1
2101	76.00	...	-1	-1
2182	53.00	...	-1	-1



jupyter project1_innomatics NEW Last Checkpoint 32 minutes ago

File Edit View Run Kernel Settings Help Not Trusted



Outliers in agreeableness:

```
    Unnamed: 0      ID  Salary      DOJ      DOL \
23     train  990077  180000 2014-08-01 2015-04-01 00:00:00
43     train  1207660  120000 2014-04-01 2015-04-01 00:00:00
63     train  636871  150000 2014-10-01 2015-04-01 00:00:00
67     train  810433  400000 2014-08-01 2015-02-01 00:00:00
157    train  728009  405000 2013-08-01           present
...
3843    train  661427  240000 2013-08-01 2013-08-01 00:00:00
3855    train  530061  300000 2014-09-01           present
3878    train  663824  260000 2013-10-01           present
3939    train  716325  100000 2013-07-01 2014-12-01 00:00:00
3953    train  331779  160000 2013-05-01           present

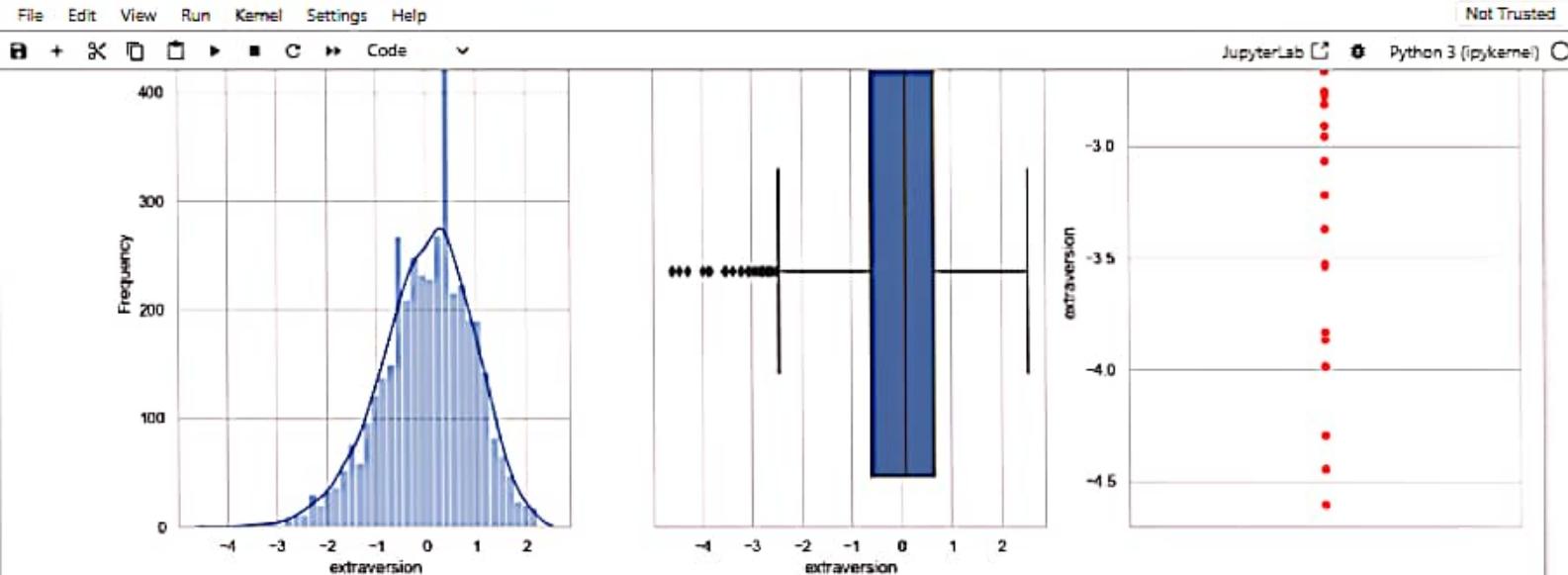
          Designation      JobCity Gender      DOB  10percentage \
23  electrical project engineer  Bangalore   m 1986-07-30    70.00
43        .net developer  Hyderabad   m 1993-07-05    59.00
63  management trainee  Bhiwadi   m 1992-01-10    68.66
67       test engineer  Bangalore   f 1991-08-25    81.00
157    system engineer  Hyderabad   m 1989-05-17    77.00
...
3843        test engineer    Jaipur   m 1991-06-11    67.17
3855  engineer trainee  Bangalore   f 1991-07-18    78.00
3878    software engineer  Bangalore   m 1989-09-10    57.70
3939    software engineer  Hyderabad   m 1992-07-05    65.00
3953  software test engineer  Chennai   m 1990-01-26    79.28

... ComputerScience  MechanicalEngg  ElectricalEngg  TelecomEngg \
23 ...
43 ...
63 ...
67 ...
157 ...
...
3843 ...
3855 ...
3878 ...
3939 ...
...
```



Scanned with OKEN Scanner

jupyter project1_innomatics NEW Last Checkpoint: 32 minutes ago



Outliers in extraversion:

	Unnamed: 0	ID	Salary	DOJ	DOL
63	train	636871	150000	2014-10-01	2015-04-01 00:00:00
159	train	755088	340000	2014-06-01	2015-04-01 00:00:00
335	train	1842343	120000	2014-02-01	present
408	train	1175511	280000	2015-02-01	2015-07-01 00:00:00
523	train	1861986	120000	2014-03-01	2015-04-01 00:00:00
666	train	820152	110000	2014-01-01	present
726	train	471411	480000	2013-07-01	2015-04-01 00:00:00
1169	train	179878	105000	2011-01-01	2011-06-01 00:00:00
1211	train	612729	420000	2013-08-01	present
1217	train	669692	600000	2014-10-01	2015-02-01 00:00:00
1242	train	319644	395000	2012-06-01	2015-04-01 00:00:00
1353	train	578644	350000	2013-06-01	present
1538	train	880117	105000	2014-12-01	2015-03-01 00:00:00
1566	train	618254	180000	2012-07-01	2014-10-01 00:00:00
1640	train	262345	375000	2011-07-01	2013-08-01 00:00:00
1728	train	955531	300000	2015-05-01	present
1785	train	133165	240000	2010-11-01	2013-08-01 00:00:00
1789	train	732482	300000	2014-05-01	2015-03-01 00:00:00
1822	train	611126	550000	2014-02-01	present
1860	train	478782	240000	2013-01-01	2015-03-01 00:00:00
2868	train	484256	115000	2012-07-01	2015-03-01 00:00:00
2224	train	624680	125000	2014-01-01	2015-04-01 00:00:00
2305	train	1058832	100000	2014-05-01	present
2377	train	618785	135000	2014-04-01	2015-04-01 00:00:00
2396	train	995921	240000	2014-01-01	2015-03-01 00:00:00
2483	train	592952	250000	2014-03-01	2014-06-01 00:00:00
3141	train	589448	450000	2010-11-01	2014-04-01 00:00:00
3158	train	1184236	145000	2014-07-01	2015-04-01 00:00:00
3174	train	248137	650000	2011-08-01	present
3202	train	538147	205000	2013-09-01	2015-06-01 00:00:00
3263	train	30768	450000	2011-01-01	2014-09-01 00:00:00
3273	train	518449	120000	2012-09-01	2013-03-01 00:00:00
3372	train	852825	310000	2014-09-01	present
3434	train	648779	210000	2013-12-01	2015-03-01 00:00:00
3473	train	1272092	35000	2014-05-01	2015-05-01 00:00:00
3548	train	681080	90000	2014-11-01	2015-05-01 00:00:00
3595	train	644505	200000	2013-12-01	2014-11-01 00:00:00
3674	train	718170	115000	2013-01-01	2013-12-01 00:00:00

jupyter project1_innomatics NEW Last Checkpoint: 32 minutes ago

159	87.00	...	-1	-1	-1	
335	68.50	...	-1	-1	324	
408	68.00	...	253	-1	-1	
523	63.00	...	-1	-1	-1	
666	59.60	...	-1	-1	-1	
726	85.00	...	-1	-1	-1	
1169	48.50	...	-1	-1	-1	
1211	80.00	...	-1	-1	-1	
1217	73.00	...	-1	313	-1	
1242	93.00	...	-1	-1	-1	
1353	85.00	...	-1	473	-1	
1538	70.06	...	223	-1	-1	
1566	79.40	...	-1	-1	-1	
1649	91.00	...	-1	-1	-1	
1728	75.00	...	-1	-1	-1	
1785	71.60	...	-1	-1	-1	
1789	78.40	...	-1	-1	-1	
1822	82.00	...	-1	-1	446	
1860	75.00	...	-1	-1	-1	
2060	69.20	...	-1	-1	-1	
2224	83.00	...	-1	-1	313	
2385	64.62	...	487	-1	-1	
2377	80.00	...	-1	-1	-1	
2396	83.40	...	-1	-1	-1	
2483	80.00	...	-1	-1	-1	
3141	85.20	...	-1	-1	-1	
3150	71.00	...	376	-1	-1	
3174	88.40	...	-1	-1	-1	
3282	90.40	...	-1	588	-1	
3263	85.66	...	-1	-1	-1	
3273	73.50	...	-1	-1	-1	
3372	86.92	...	-1	-1	-1	
3434	64.00	...	-1	-1	-1	
3473	85.30	...	376	-1	-1	
3548	67.00	...	-1	-1	-1	
3595	91.60	...	-1	-1	-1	
3674	52.00	...	-1	-1	-1	
3694	72.00	...	-1	-1	-1	
3778	87.20	...	346	-1	-1	
	TelecomEngg	CivilEngg	conscientiousness	agreeableness	extraversion	\
63	-1	-1	-2.2351	-2.6847	-2.6028	
159	-1	-1	-2.8879	-4.2831	-3.2176	
335	-1	-1	-3.6860	-5.6151	-4.6009	
408	-1	-1	-3.3188	-1.6191	-3.5250	
523	-1	-1	-3.1752	-5.1156	-3.0639	
666	-1	-1	0.4155	-1.2861	-3.2176	
726	-1	-1	-1.9629	-2.1903	-2.7750	
1169	-1	-1	-2.1175	-3.0874	-2.6662	
1211	-1	-1	-3.4624	-5.7815	-4.6009	
1217	-1	322	-2.0262	-1.9521	-2.6028	
1242	-1	-1	-8.8810	1.1248	-2.9565	
1353	-1	-1	-3.1752	-4.2831	-4.2935	
1538	-1	-1	-2.7443	-2.4516	-3.0639	
1566	-1	-1	-2.0262	-3.7836	-2.7565	
1649	473	-1	-2.1175	-1.8393	-1.5370	
1728	473	-1	-0.3027	-0.6201	-3.5250	
1785	-1	-1	-2.1175	-1.3713	-2.6662	
1789	-1	-1	0.4155	0.5454	-2.7565	
1822	-1	-1	-0.1590	-1.2861	-2.7565	
1860	-1	-1	-2.2720	-3.1264	-3.8636	
2060	-1	-1	-1.3080	-1.1196	-3.0639	
2224	-1	-1	-2.8879	-3.9501	-2.7565	

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

488	0.27278	-0.4776
523	-0.74158	-5.2679
666	-1.75568	-3.3518
726	0.41488	-0.4455
1169	0.06238	-3.4471
1211	-0.36120	-6.9925
1217	0.01920	-2.5853
1242	-0.78967	-1.2354
1353	-2.13688	-5.8428
1538	-0.48798	-3.5434
1566	0.14598	-6.8089
1649	2.17748	-0.4455
1728	-0.23448	-1.4356
1785	1.47248	0.0284
1789	-1.62898	0.2889
1822	0.81928	-2.5853
1860	0.29738	-3.6851
2860	-0.36128	-0.6692
2224	-1.24858	-5.6512
2385	1.79488	1.0554
2377	-1.12188	-5.4595
2396	-0.48798	-5.8428
2483	0.27278	0.4885
3141	0.76738	0.1864
3150	-0.86828	-6.6892
3174	-0.76838	0.3444
3282	1.07478	-2.0253
3263	0.06238	0.0284
3273	-0.26098	0.8183
3372	-0.74158	-5.2679
3434	-1.24858	-1.6273
3473	-0.48798	-7.3757
3548	0.39958	-1.6273
3595	-0.86828	0.0973
3674	1.03338	-2.7769
3694	-0.74158	-6.9925
3778	0.77988	0.2889

[48 rows x 39 columns]

SHOW MORE OUTPUTS

```
[19]: # Check unique values in the categorical columns
for col in categorical_cols:
    print(f"Unique values in {col}:")
    print(df[col].unique())
    print("\n")
```

```
Unique values in Unnamed: 0:
['train']
```

```
Unique values in DOL:
['present' datetime.datetime(2015, 3, 1, 0, 0)
 datetime.datetime(2015, 5, 1, 0, 0) datetime.datetime(2015, 7, 1, 0, 0)
 datetime.datetime(2015, 4, 1, 0, 0) datetime.datetime(2014, 10, 1, 0, 0)
 datetime.datetime(2014, 9, 1, 0, 0) datetime.datetime(2014, 6, 1, 0, 0)
 datetime.datetime(2012, 9, 1, 0, 0) datetime.datetime(2013, 12, 1, 0, 0)
 datetime.datetime(2015, 6, 1, 0, 0) datetime.datetime(2013, 10, 1, 0, 0)
 datetime.datetime(2015, 1, 1, 0, 0) datetime.datetime(2014, 4, 1, 0, 0)]
```



Scanned with OKEN Scanner

Project 1 - Innomatics

Jupyter project1_innomatics NEW Last Checkpoint 32 minutes ago



File Edit View Run Kernel Settings Help

Not Trusted

Code

JupyterLab Python 3 (ipykernel)

```
'senior software engineer' 'get' 'system engineer'  
'java software engineer' 'mechanical engineer' 'electrical engineer'  
'project engineer' 'senior php developer' 'senior systems engineer'  
'quality assurance engineer' 'qa analyst' 'network engineer'  
'product development engineer' 'associate software developer'  
'data entry operator' 'software engineer' 'developer'  
'electrical project engineer' 'programmer analyst' 'systems analyst'  
'ase' 'telecommunication engineer' 'application developer'  
'ios developer' 'executive assistant' 'online marketing manager'  
'documentation specialist' 'associate software engineer'  
'management trainee' 'site manager' 'software developer' '.net developer'  
'production engineer' 'jr. software engineer'  
'trainee software developer' 'ui developer' 'assistant system engineer'  
'android developer' 'customer service' 'test engineer' 'java developer'  
'engineer' 'recruitment coordinator' 'technical support engineer'  
'data analyst' 'assistant software engineer' 'faculty'  
'entry level management trainee' 'customer service representative'  
'software test engineer' 'firmware engineer' 'php developer'  
'research associate' 'research analyst' 'quality engineer' 'programmer'  
'technical support executive' 'business analyst' 'web developer'  
'application engineer' 'project coordinator' 'engineer trainee'  
'sap consultant' 'quality analyst' 'marketing coordinator'  
'system administrator' 'senior engineer' 'business development managerde'  
'network administrator' 'technical support specialist'  
'business development executive' 'junior software engineer'  
'asp.net developer' 'graduate engineer trainee' 'field engineer'  
'assistant professor' 'trainee software engineer'  
'senior software developer' 'quality assurance automation engineer'  
'design engineer' 'telecom engineer' 'quality control engineer'  
'hardware engineer' 'hr recruiter' 'sales associate' 'junior engineer'  
'associate engineer' 'maintenance engineer' 'sales engineer'  
'human resources associate' 'mobile application developer'  
'electronic field service engineer' 'process associate'  
'field service engineer' 'it support specialist'  
'software development engineer' 'business process analyst'  
'operation engineer' 'electrical designer' 'marketing assistant'  
'sales executive' 'admin assistant' 'senior java developer'  
'account executive' 'oracle dba' 'rf engineer'  
'embedded software engineer' 'programmer analyst trainee'  
'technical engineer' 'operations executive' 'trainee engineer'  
'recruiter' 'lecturer' '.net web developer' 'marketing executive'  
'operations assistant' 'associate manager' 'electrical design engineer'  
'systems administrator' 'client services associate' 'it analyst'  
'senior developer' 'cad designer' 'business technology analyst'  
'asst. manager' 'service engineer' 'executive recruiter'  
'planning engineer' 'associate technical operations' 'web designer'  
'software architect' 'software quality assurance tester' 'seo trainee'  
'process engineer' 'software quality assurance analyst' 'designer'  
'business systems consultant' 'business development manager'  
'junior research fellow' 'technical recruiter' 'operations analyst'  
'quality assurance test engineer' 'linux systems administrator'  
'software trainee' 'entry level sales and marketing'  
'electrical field engineer' 'windows systems administrator'  
'junior software developer' 'python developer'  
'web application developer' 'assistant systems engineer'  
'javascript developer' 'operation executive' 'performance engineer'  
'technical writer' 'operations engineer and jetty handling'  
'lead engineer' 'portfolio analyst' 'associate system engineer'  
'mechanical design engineer' 'product engineer'  
'network security engineer' 'operations manager' 'technical lead'  
'operations' 'quality assurance tester' 'automation engineer'  
'data scientist' 'quality associate' 'manual tester' 'sr. engineer'  
'embedded engineer' 'service and sales engineer'  
'internet support engineer' 'engineer- customer support' 'cloud engineer'
```



Scanned with OKEN Scanner

```
'network security engineer' 'operations manager' 'technical lead'  
'operations' 'quality assurance tester' 'automation engineer'  
'data scientist' 'quality associate' 'manual tester' 'sr. engineer'  
'embedded engineer' 'service and sales engineer'  
'telecom support engineer' 'engineer- customer support' 'cloud engineer'  
'branch manager' 'business analyst consultant' 'technology lead'  
'software trainee engineer' 'dcs engineer' 'junior manager' 'ux designer'  
'clerical' 'hr generalist' 'database administrator'  
'senior design engineer' 'seo' 'assistant engineer' 'marketing analyst'  
'it executive' 'salesforce developer' 'software tester' 'sql dba'  
'junior engineer product support' 'manager' 'senior business analyst'  
'c# developer' 'implementation engineer' 'executive hr'  
'executive engineer' 'sharepoint developer' 'system analyst'  
'sales management trainee' 'senior project engineer' 'it recruiter'  
'software engineer analyst' 'desktop support technician'  
'continuous improvement engineer' 'process advisor' 'etl developer'  
'sales and service engineer' 'project manager' 'training specialist'  
'product manager' 'staffing recruiter' 'assistant programmer'  
'quality controller' 'mis executive' 'game developer'  
'digital marketing specialist' 'principal software engineer'  
'software developer' 'senior mechanical engineer'  
'technical operations analyst' 'service coordinator' 'testing engineer'  
'technical assistant' 'sap abap consultant' 'seo engineer'  
'project assistant' 'talent acquisition specialist'  
'sales account manager' 'software engineer trainee'  
'customer service manager' 'help desk analyst' 'general manager'  
'engineering manager' 'senior network engineer'  
'field based employee relations manager' 'phone banking officer'  
'support engineer' 'associate test engineer' 'technology analyst'  
'network support engineer' 'it business analyst' 'junior system analyst'  
'senior .net developer' 'secretary' 'research engineer'  
'quality assurance auditor' 'process executive'  
'lecturer & electrical maintenance' 'office coordinator' 'hr manager'  
'html developer' 'sales support' 'front end web developer'  
'administrative support' 'territory sales manager'  
'project administrator' 'environmental engineer' 'web designer and seo'  
'information security analyst' 'field business development associate'  
'operational executive' 'administrative coordinator'  
'senior risk consultant' 'desktop support engineer' 'cad drafter'  
'not engineer' 'industrial engineer' 'it engineer'  
'human resources intern' 'senior quality assurance engineer'  
'clerical assistant' 'software engineer' 'quality assurance'  
'delivery software engineer' 'graphic designer'  
'sales development manager' 'visiting faculty'  
'business intelligence analyst' 'team lead'  
'operational excellence manager' 'sales & service engineer' 'web intern'  
'full stack developer' 'database developer' 'sr. database engineer'  
'graduate apprentice trainee' 'software engineer associate'  
'technical analyst' 'executive engg' 'it technician'  
'business system analyst' 'process control engineer'  
'technical consultant' 'business office manager'  
'quality control inspector' 'product design engineer'  
'manufacturing engineer' 'seo executive' 'sap analyst'  
'software engineer' 'financial service consultant' 'co faculty'  
'software analyst' 'desktop support analyst' 'graduate engineer'  
'engineering technician' 'it assistant' 'marketing manager'  
'human resource assistant' 'hr assistant' 'product developer'  
'customer support engineer' 'quality control inspection technician'  
'gis/cad engineer' 'senior web developer' 'sql developer'  
'research staff member' 'sap abap associate consultant' 'associate qa'  
'corporate recruiter' 'project management officer'  
'business systems analyst' 'software programmer' 'help desk technician'  
'sales manager' 'catalog associate' 'assistant store manager'
```

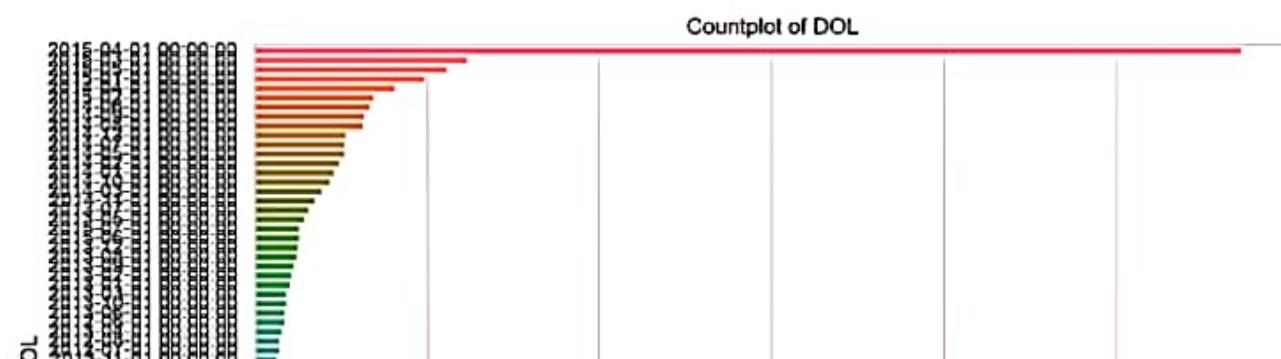
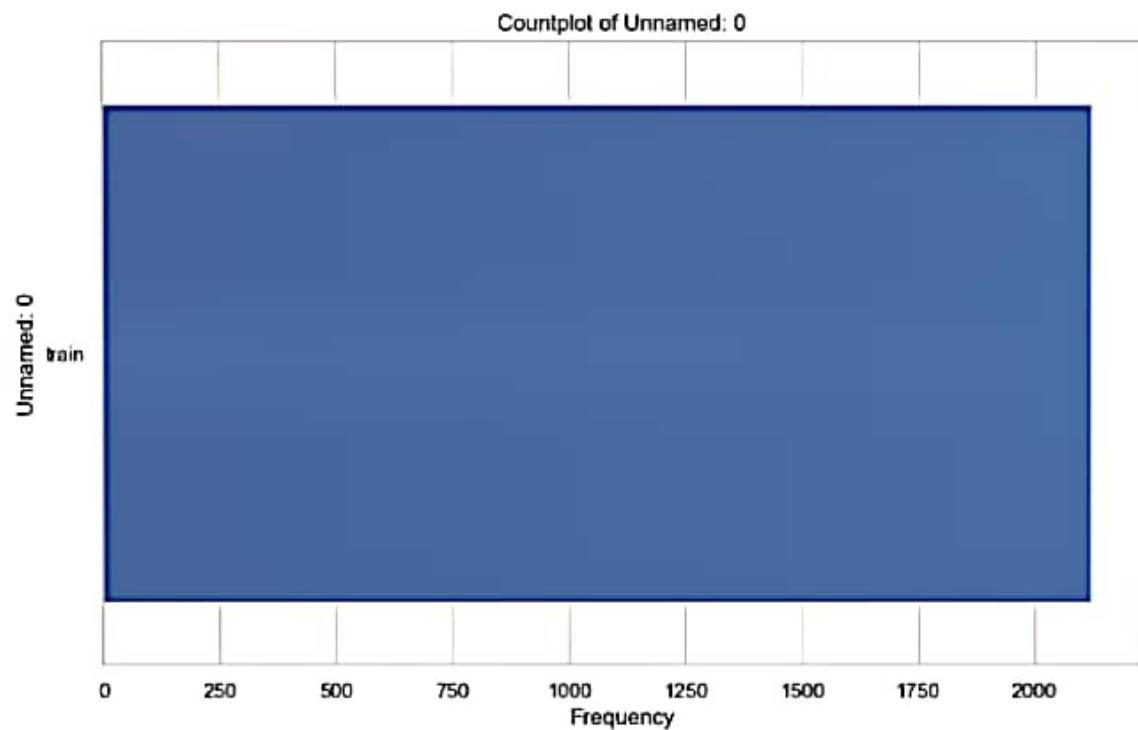
jupyter project1_innomatics NEW Last Checkpoint 33 minutes ago

File Edit View Run Kernel Settings Help

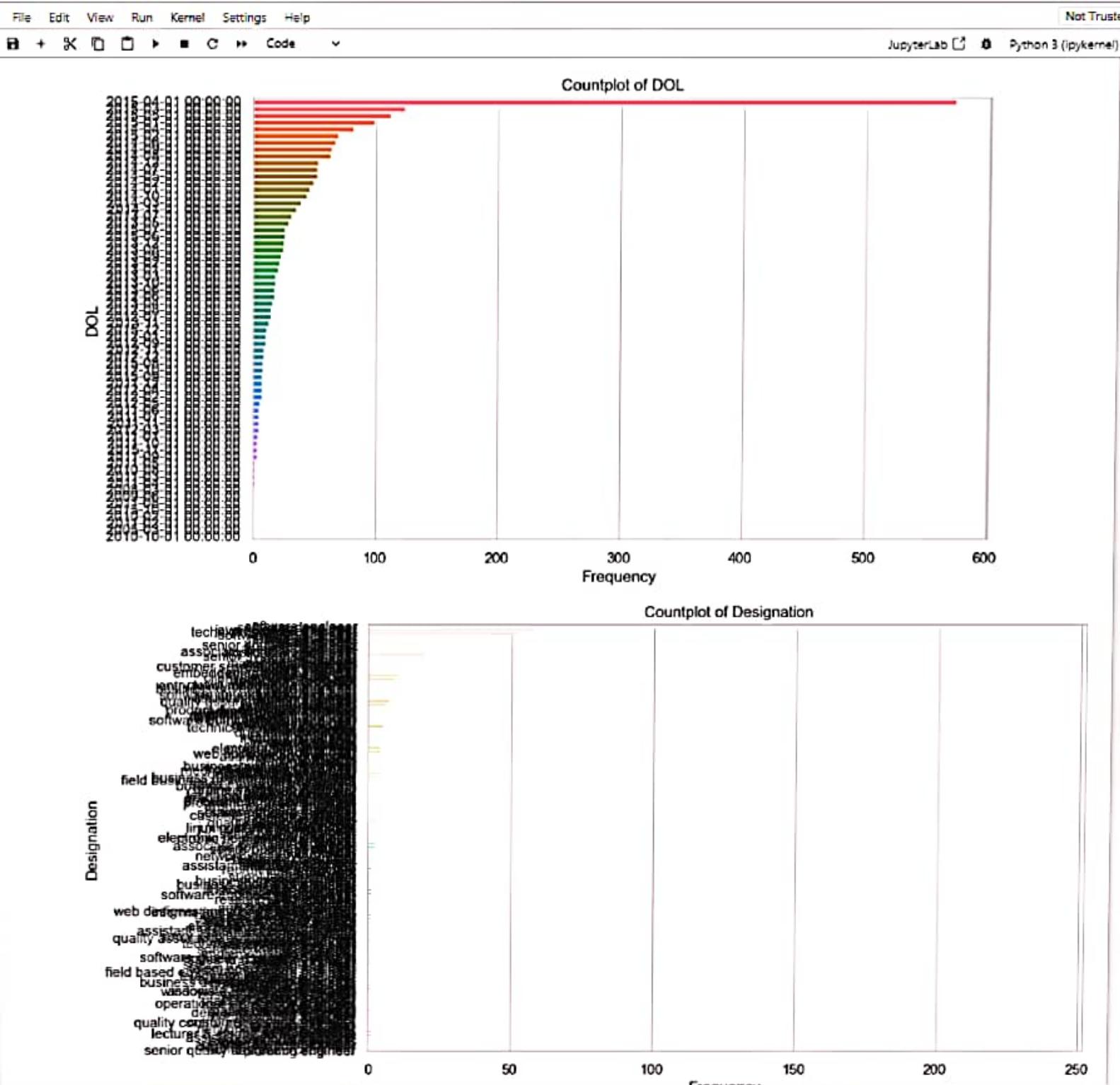
Not Trusted

```
[20]: # Drop rows where 'DOL' has invalid values
df = df[df['DOL'] != 'present'] # Replace 'present' with an appropriate action as necessary

[21]: # Recreate count plots for categorical variables
for col in categorical_cols:
    plt.figure(figsize=(10, 6))
    sns.countplot(y=df[col], order=df[col].value_counts().index)
    plt.title(f'Countplot of {col}')
    plt.xlabel('Frequency')
    plt.ylabel(col)
    plt.show()
```



jupyter project1_innomatics NEW Last Checkpoint: 33 minutes ago



jupyter project1_innomatics NEW Last Checkpoint 33 minutes ago



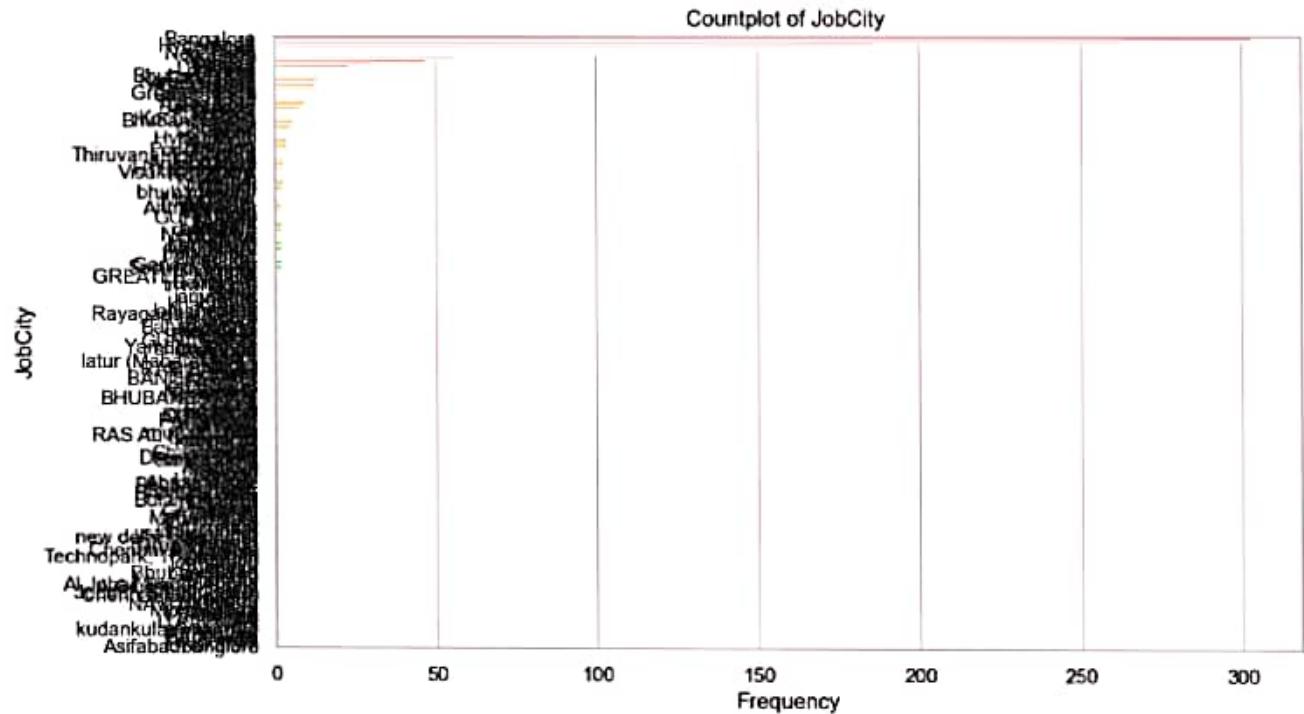
File Edit View Run Kernel Settings Help

Not Trusted

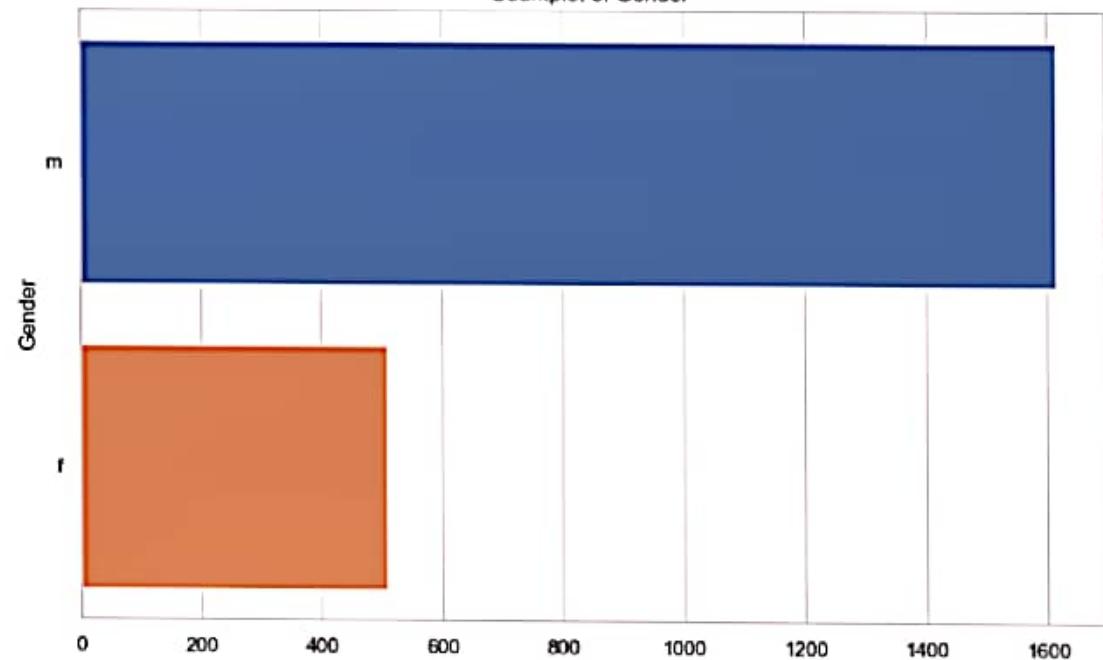
Code

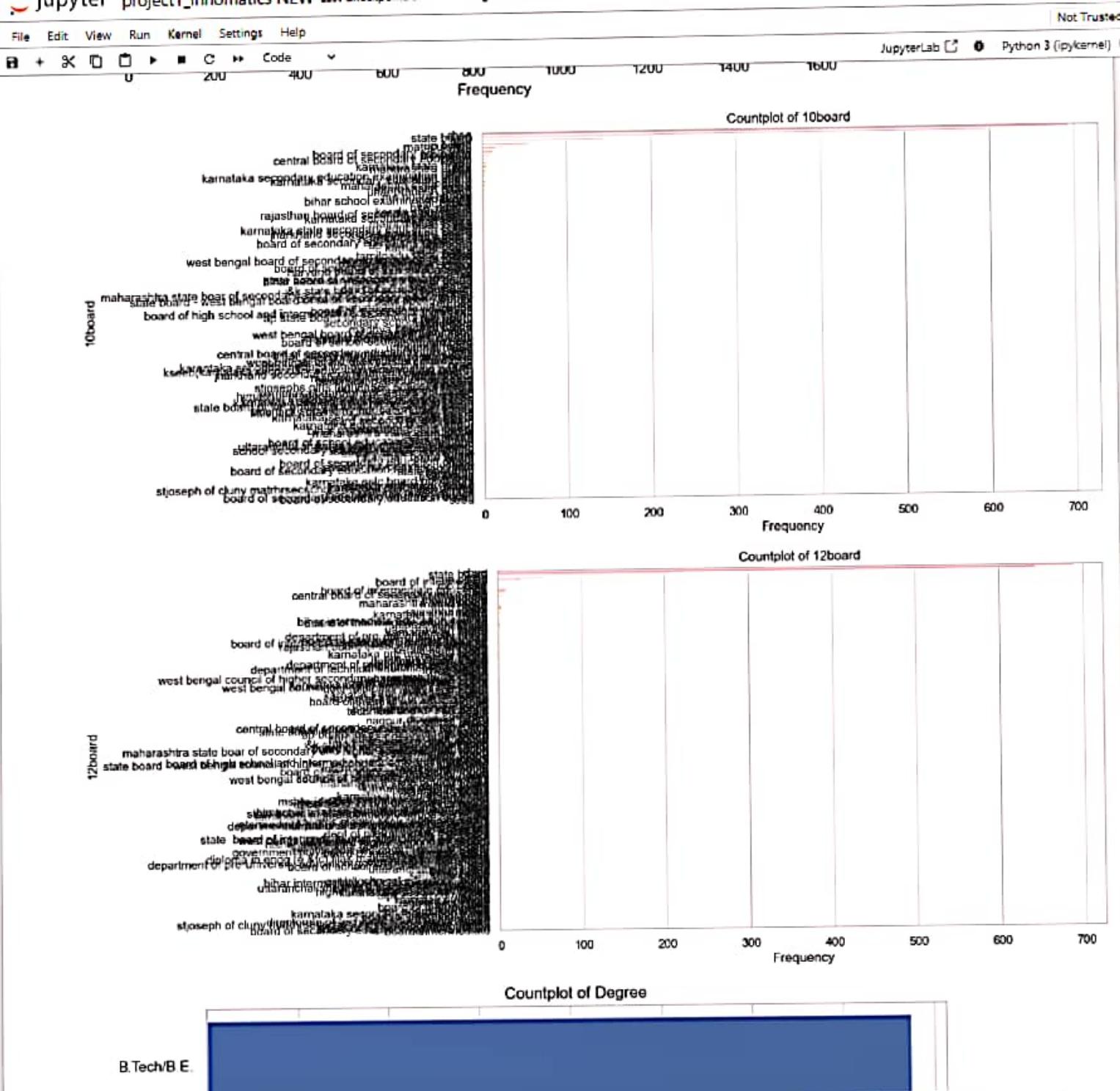
JupyterLab

Python 3 (ipykernel)



Countplot of Gender





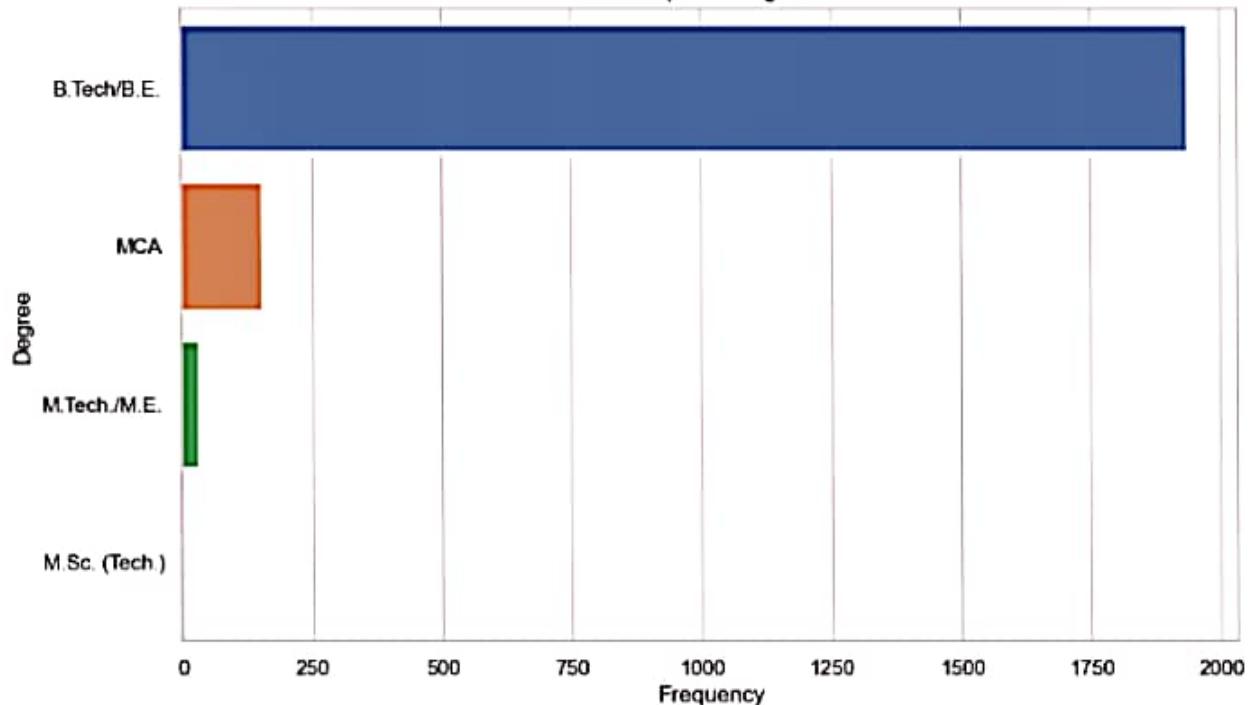
jupyter project1_innomatics NEW Last Checkpoint 33 minutes ago

File Edit View Run Kernel Settings Help Not Trusted

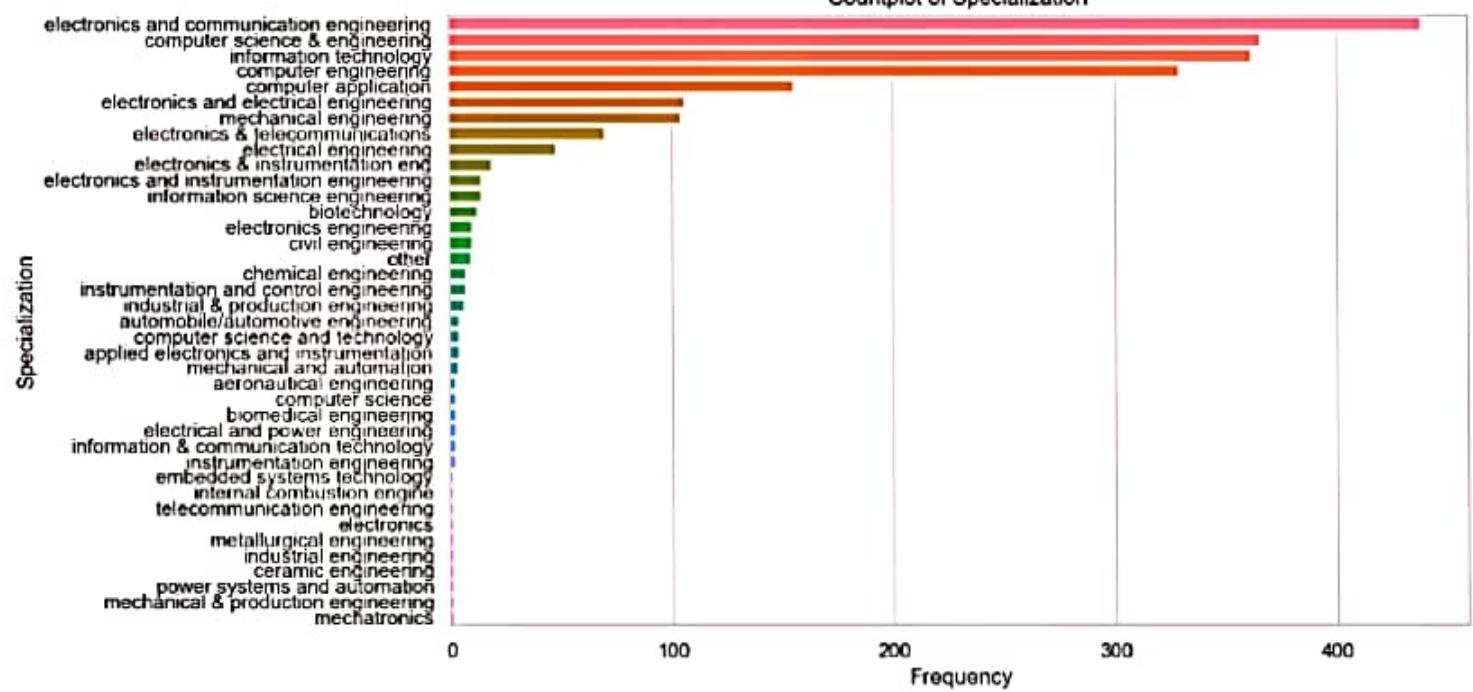
Code

JupyterLab Python 3 (ipykernel)

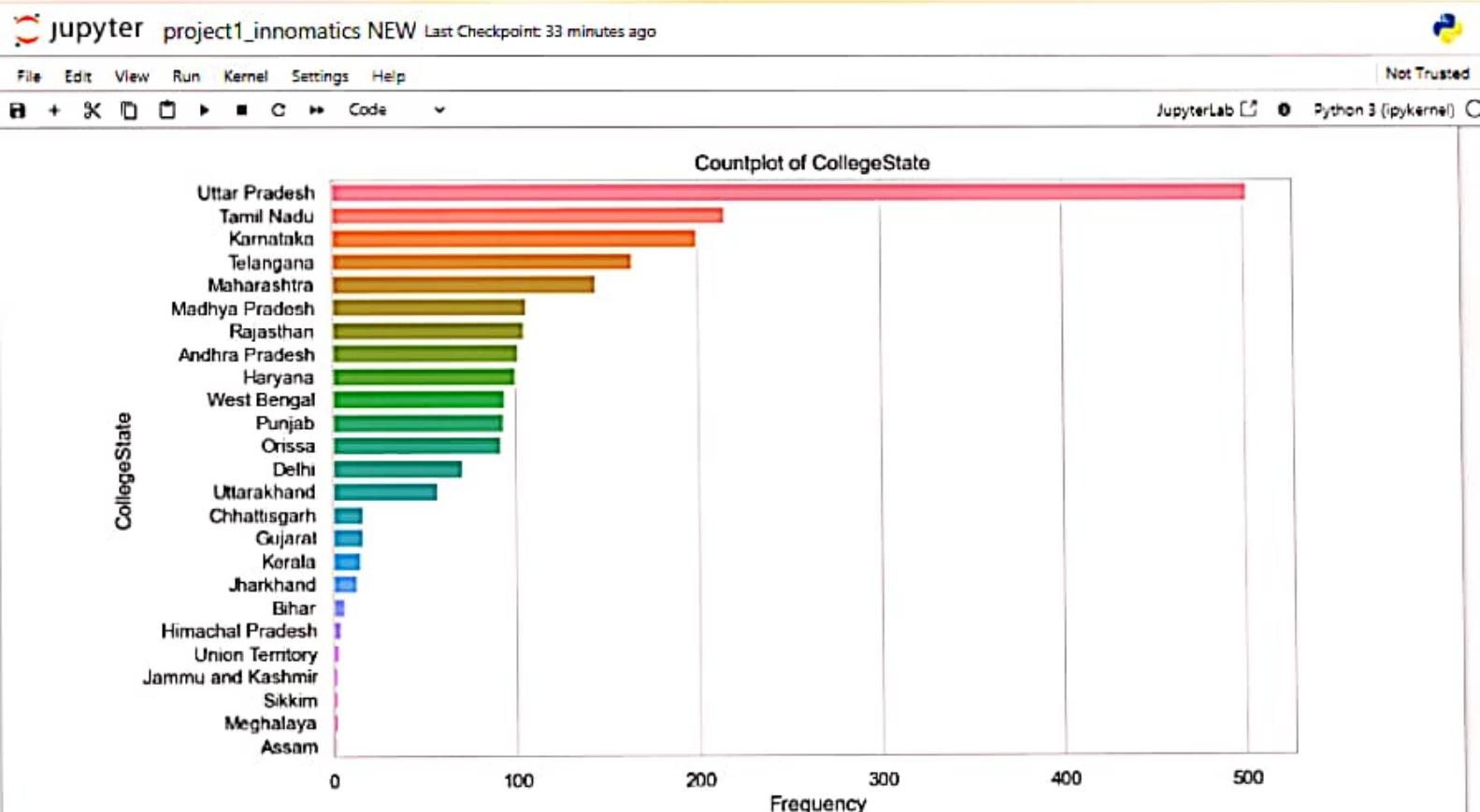
Countplot of Degree



Countplot of Specialization



NOMATICS%20OCTOBER/project1_innomatics%20NEW.ipynb



Observations:

Numerical Variables: Salary:

Histogram: Shows a right-skewed distribution. Many graduates earn between ₹2,00,000 and ₹4,00,000, with a few outliers earning significantly more. Boxplot: Indicates potential outliers above the upper quartile. The range of salaries is quite wide, suggesting variability in salaries based on various factors. 10percentage, 12percentage:

Histogram: Both show a normal-like distribution, with most students scoring between 60% and 90%. Boxplot: A few outliers exist, particularly for those scoring exceptionally low or high. College GPA:

Histogram: Indicates a slightly right-skewed distribution, with most students having GPAs between 6 and 8. Boxplot: Presence of outliers, particularly on the higher side, which could indicate exceptionally high achievers. Cognitive Scores (English, Logical, Quant):

Histograms: All three distributions appear slightly skewed, with most scores clustering towards the higher end. Boxplots: Outliers detected, particularly for logical and quantitative scores, indicating potential exceptional performances. Personality Traits:

Histograms: Mostly clustered around the mean for all traits, with slight deviations. Boxplots: A few extreme values exist, indicating candidates with unusual scores. Categorical Variables: Gender:

Countplot: Shows a relatively balanced distribution between male and female candidates. Designation:

Countplot: Some designations (like 'Software Engineer') appear more frequently than others, indicating popular job roles. JobCity:

Countplot: Cities like Bangalore and Hyderabad have significantly more job placements, reflecting industry hubs in the engineering sector. College Tier:

Step 4: Bivariate Analysis

4.1. Numerical vs. Numerical Relationships Scatter Plot

You can use scatter plots to visualize relationships between two numerical variables. For example, you can analyze the relationship between Salary and CollegeGPA.

```
[24]: print(df.columns)
```

```
Index(['Unnamed: 0', 'ID', 'Salary', 'DOJ', 'DOL', 'Designation', 'JobCity',
       'Gender', 'DOB', '10percentage', '10board', '12graduation',
       '12percentage', '12board', 'CollegeID', 'CollegeTier', 'Degree',
       'Specialization', 'collegeGPA', 'CollegeCityID', 'CollegeCityTier',
       'CollegeState', 'GraduationYear', 'English', 'Logical', 'Quant',
       'Domain', 'ComputerProgramming', 'ElectronicsAndSemicon',
       'ComputerScience', 'MechanicalEngg', 'ElectricalEngg', 'TelecomEngg',
       'CivilEngg', 'conscientiousness', 'agreeableness', 'extraversion',
       'nueroticism', 'openness_to_experience'],
      dtype='object')
```

```
[25]: print(df.dtypes)
```

```
Unnamed: 0          object
ID              int64
Salary          int64
DOJ            datetime64[ns]
DOL            object
Designation    object
JobCity        object
Gender          object
DOB            datetime64[ns]
10percentage   float64
10board         object
12graduation   int64
12percentage   float64
12board         object
CollegeID      int64
CollegeTier    int64
Degree          object
Specialization object
collegeGPA     float64
CollegeCityID  int64
CollegeCityTier int64
CollegeState    object
GraduationYear int64
English         int64
Logical         int64
Quant           int64
Domain          float64
ComputerProgramming int64
ElectronicsAndSemicon int64
ComputerScience   int64
MechanicalEngg    int64
ElectricalEngg     int64
TelecomEngg       int64
CivilEngg        int64
conscientiousness float64
agreeableness    float64
extraversion      float64
nueroticism      float64
```

jupyter project1_innomatics NEW Last Checkpoint 33 minutes ago

File Edit View Run Kernel Settings Help

Not Trusted

dtype: object

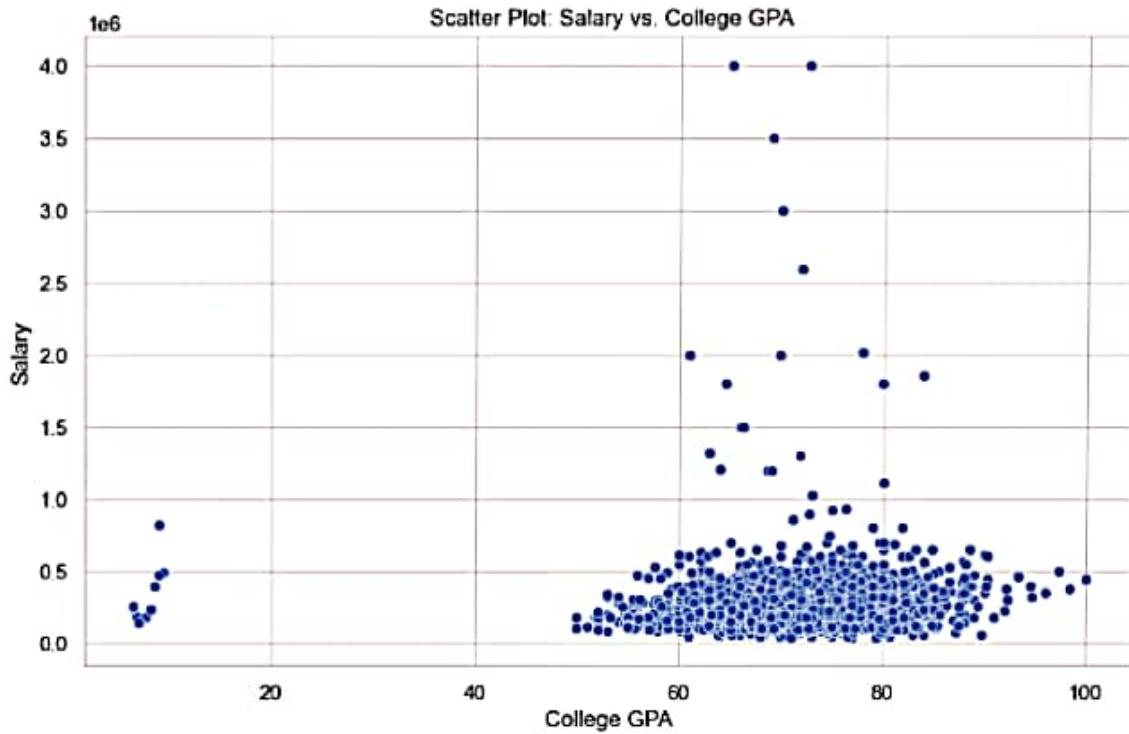
JupyterLab

Python 3 (ipykernel)

```
[26]: plt.figure(figsize=(10, 6))
sns.scatterplot(x='collegeGPA', y='Salary', data=df) # Use 'collegeGPA' instead of 'CollegeGPA'
plt.title('Scatter Plot: Salary vs. College GPA')
plt.xlabel('College GPA')
plt.ylabel('Salary')
plt.show()
```

Observations:

Look for trends, such as whether higher GPAs correlate with higher salaries.
Identify any clusters or outliers.



```
[28]: # Hexbin Plot
```

```
# For large datasets, hexbin plots are effective for visualizing the density of points.
# Hexbin Plot
# For large datasets, hexbin plots are effective for visualizing the density of points.
plt.figure(figsize=(10, 6))
plt.hexbin(df['collegeGPA'], df['Salary'], gridsize=50, cmap='Blues') # Use 'collegeGPA'
plt.colorbar(label='Count in Hexbin')
plt.title('Hexbin Plot: Salary vs. College GPA')
plt.xlabel('College GPA')
plt.ylabel('Salary')
plt.show()
```



Scanned with OKEN Scanner

jupyter project1_innomatics NEW Last Checkpoint 33 minutes ago

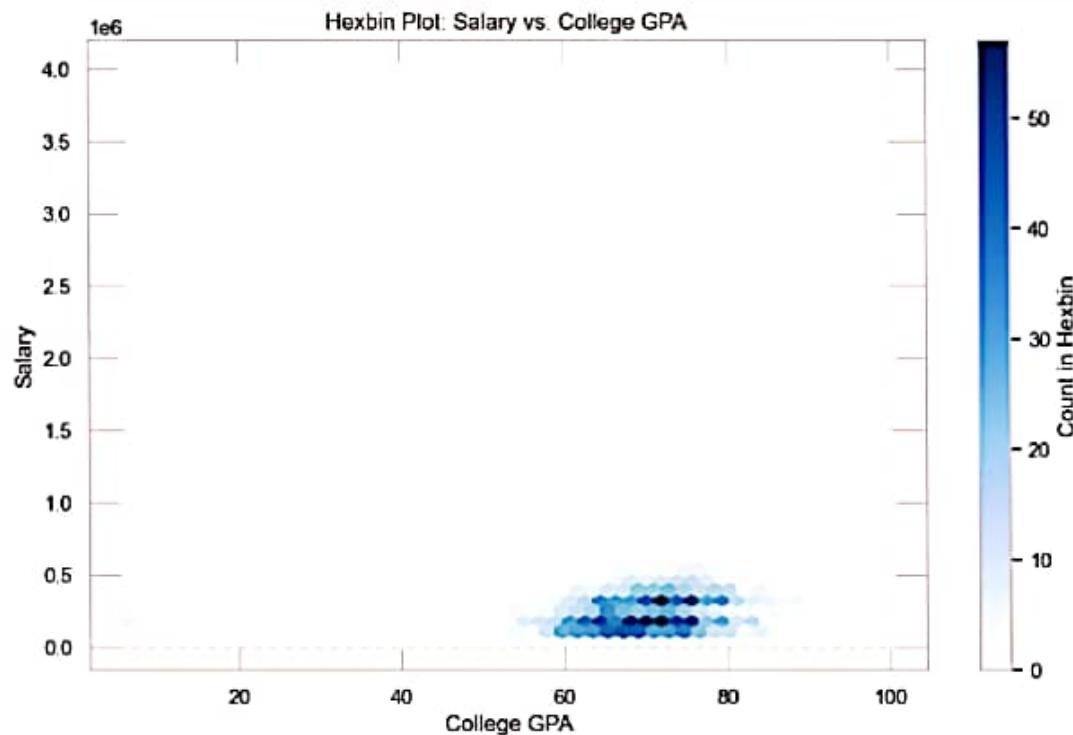
File Edit View Run Kernel Settings Help Not Trusted

Code ▾

JupyterLab Python 3 (ipykernel)

[28]: # Hexbin Plot

```
# For large datasets, hexbin plots are effective for visualizing the density of points.  
# Hexbin Plot  
# For large datasets, hexbin plots are effective for visualizing the density of points.  
plt.figure(figsize=(10, 6))  
plt.hexbin(df['collegeGPA'], df['Salary'], gridsize=50, cmap='Blues') # Use 'collegeGPA'  
plt.colorbar(label='Count in Hexbin')  
plt.title('Hexbin Plot: Salary vs. College GPA')  
plt.xlabel('College GPA')  
plt.ylabel('Salary')  
plt.show()  
  
# Observations:  
  
# Assess the density of data points in different regions.  
# Determine if higher densities correspond with higher salaries.
```



[30]: # Pair Plot

```
# A pair plot allows you to visualize relationships across multiple numerical columns.  
sns.pairplot(df[['Salary', 'collegeGPA', '10percentage', '12percentage']]) # Use 'collegeGPA'  
plt.suptitle('Pair Plot of Salary, College GPA, and Percentages', y=1.02)  
plt.show()  
  
# Observations:
```

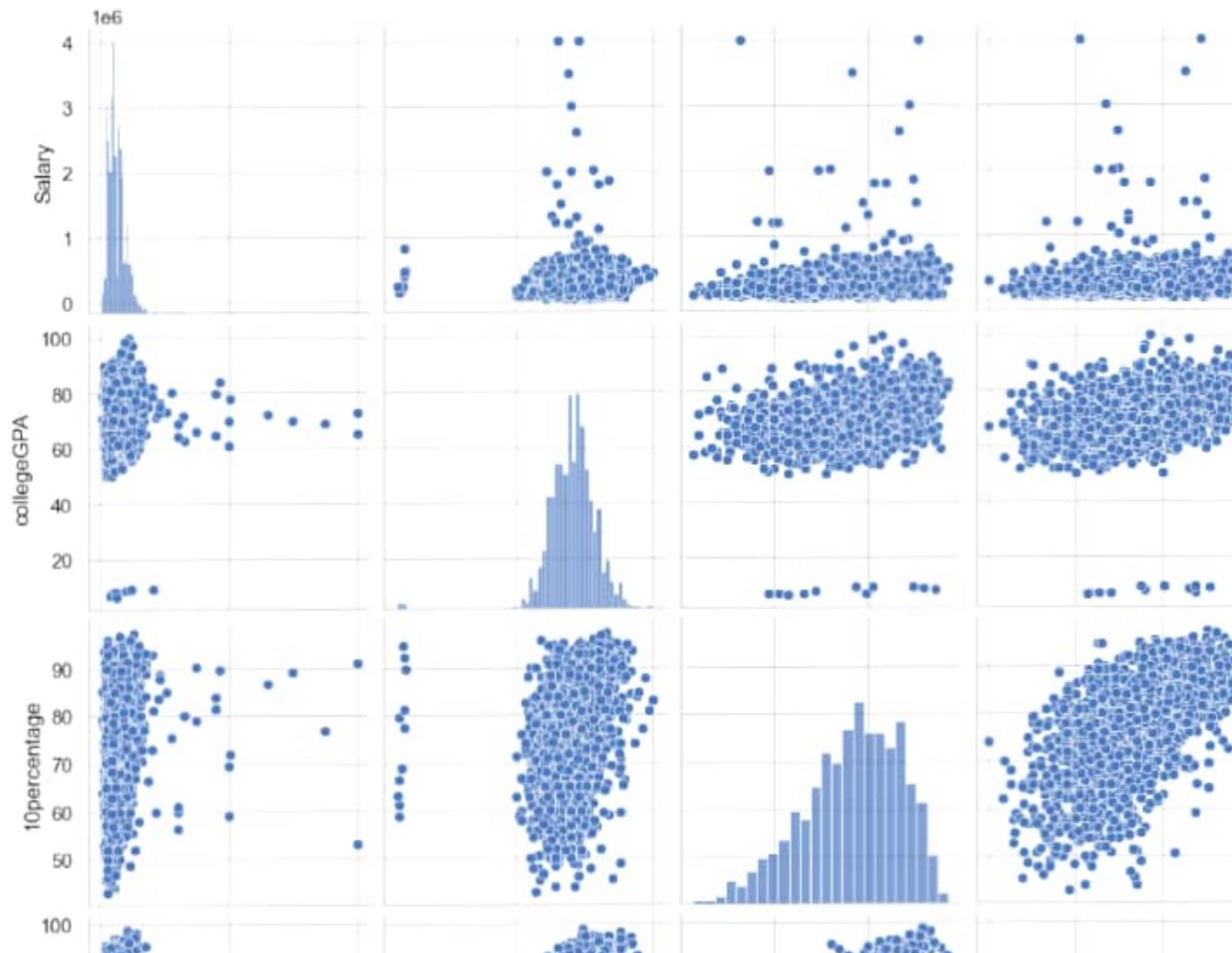
```
[58]: # Pair Plot  
  
# A pair plot allows you to visualize relationships across multiple numerical columns.  
sns.pairplot(df[['Salary', 'collegeGPA', '10percentage', '12percentage']]) # Use 'collegeGPA'  
plt.suptitle('Pair Plot of Salary, College GPA, and Percentages', y=1.02)  
plt.show()
```

Observations:

Examine the pairwise relationships and distributions.
Look for any strong linear relationships.

```
C:\Users\padal\anaconda3\lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight  
self._figure.tight_layout(*args, **kwargs)
```

Pair Plot of Salary, College GPA, and Percentages



Scanned with OKEN Scanner

File Edit View Run Kernel Settings Help

Not Trusted

Salary 100 collegeGPA 10percentage

JupyterLab Python 3 (ipykernel) 12percentage

```
[31]: # Categorical vs. Numerical Relationships  
# Box Plot
```

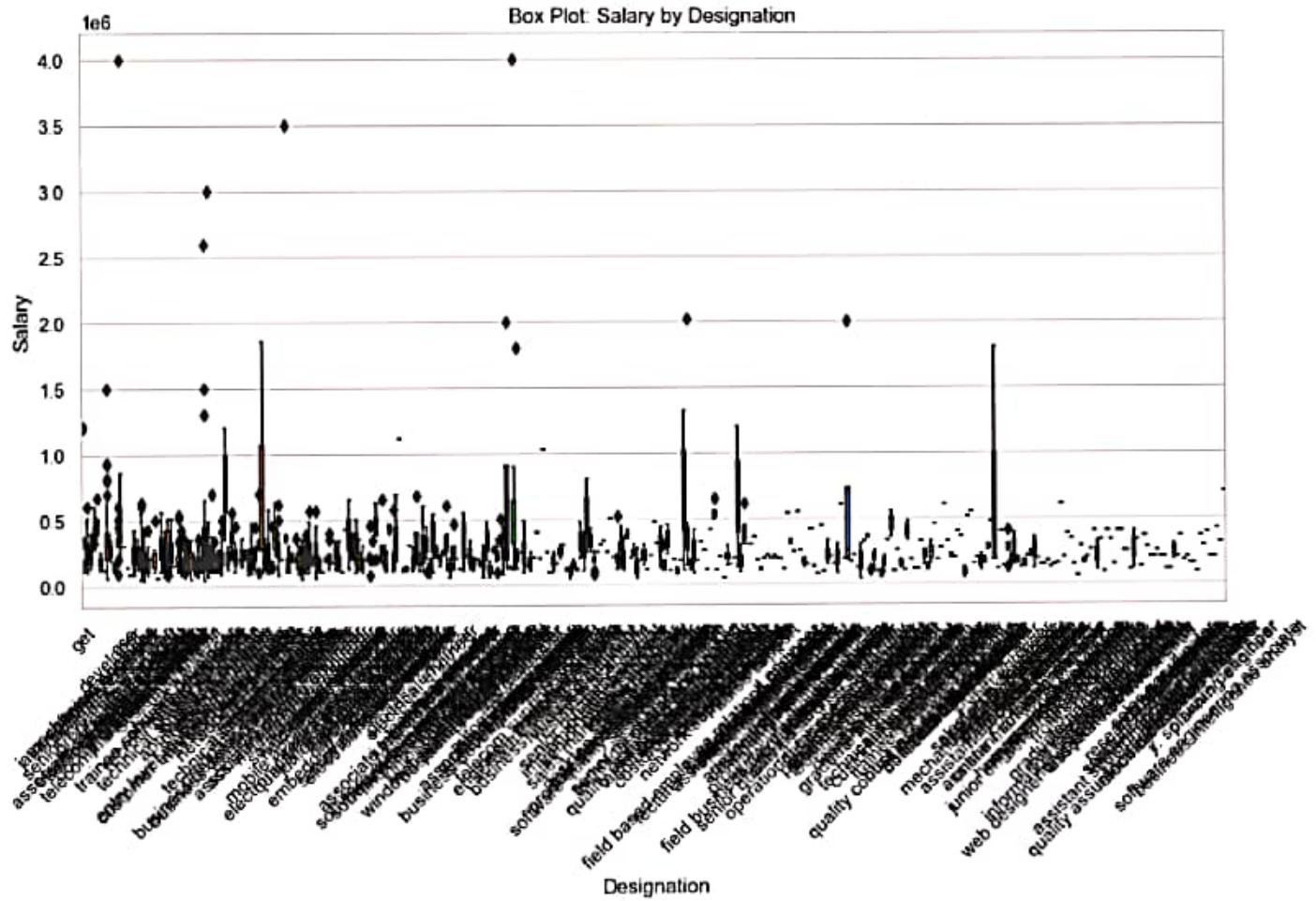
Box plots are useful for visualizing the distribution of salaries across different designations.

```
plt.figure(figsize=(12, 6))
sns.boxplot(x='Designation', y='Salary', data=df)
plt.title('Box Plot: Salary by Designation')
plt.xticks(rotation=45)
plt.show()
```

Observations:

Assess the median and spread of salaries for each designation.

Identify any outliers within the salary distribution for different roles.



File Edit View Run Kernel Settings Help Not T

jupyter project1_innomatics NEW Last Checkpoint: 34 minutes ago

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

Designation

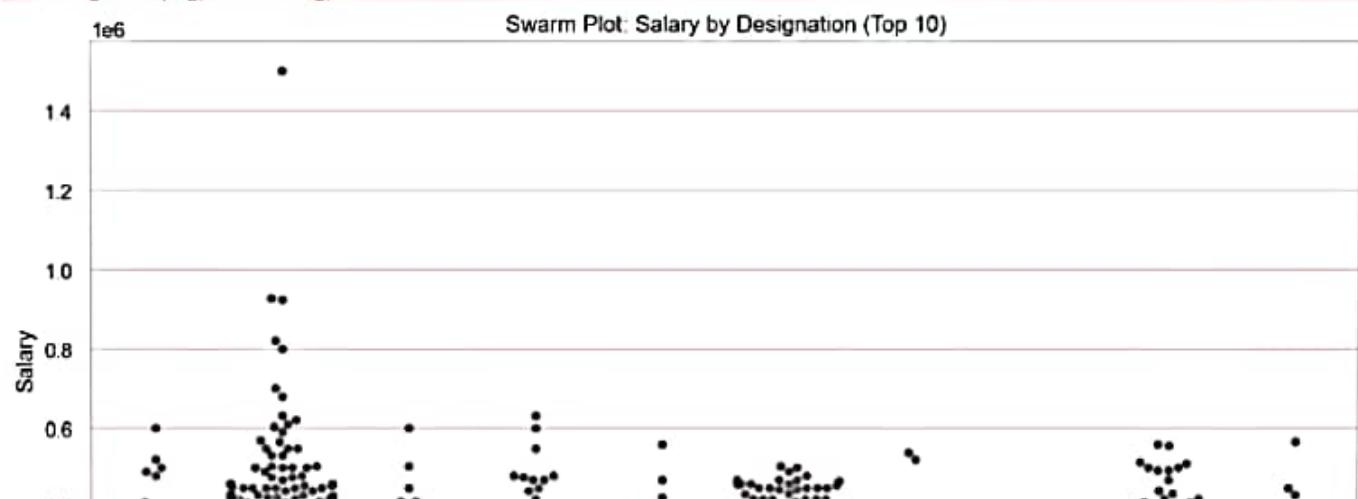
[33]: # Swarm Plot

```
# Swarm plots can complement box plots by showing individual data points.  
# Limit swarm plot to top 10 most frequent Designations  
top_designations = df['Designation'].value_counts().nlargest(10).index  
filtered_df = df[df['Designation'].isin(top_designations)]  
  
plt.figure(figsize=(12, 6))  
sns.swarmplot(x='Designation', y='Salary', data=filtered_df, color='black', alpha=0.5)  
plt.title('Swarm Plot: Salary by Designation (Top 10)')  
plt.xticks(rotation=45)  
plt.show()
```

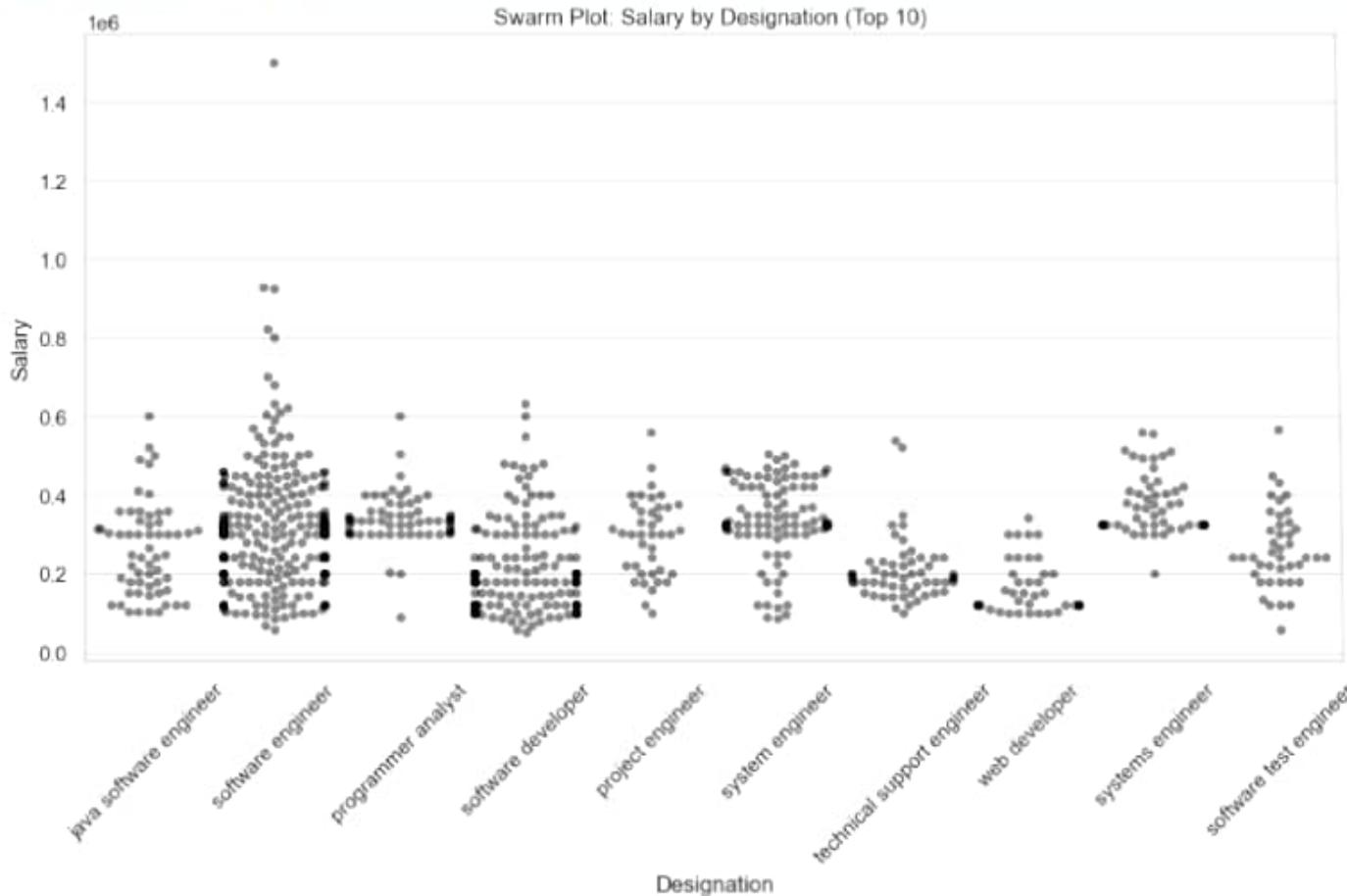
Observations:

```
# Visualize the actual salary data points along with the box plot.  
# Identify any overlaps or clusters in salary distributions for specific designations.
```

```
C:\Users\padal\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning: 34.2% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.  
warnings.warn(msg, UserWarning)  
C:\Users\padal\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning: 21.1% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.  
warnings.warn(msg, UserWarning)  
C:\Users\padal\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning: 17.0% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.  
warnings.warn(msg, UserWarning)  
C:\Users\padal\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning: 14.5% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.  
warnings.warn(msg, UserWarning)  
C:\Users\padal\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning: 15.4% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.  
warnings.warn(msg, UserWarning)  
C:\Users\padal\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning: 17.3% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.  
warnings.warn(msg, UserWarning)
```



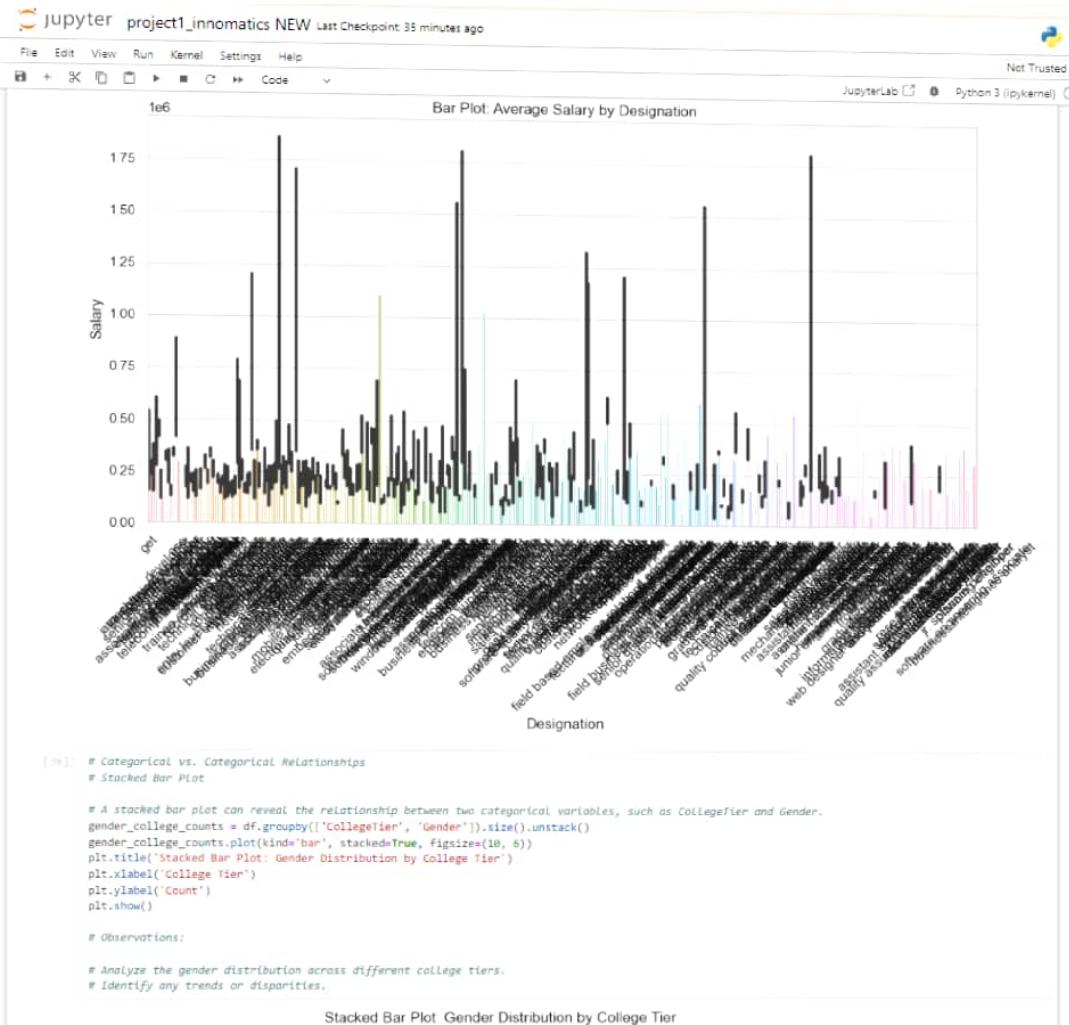
```
warnings.warn(msg, UserWarning)
C:\Users\padal\anaconda3\Lib\site-packages\seaborn\categorical.py:3544: UserWarning: 17.3% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
warnings.warn(msg, UserWarning)
```



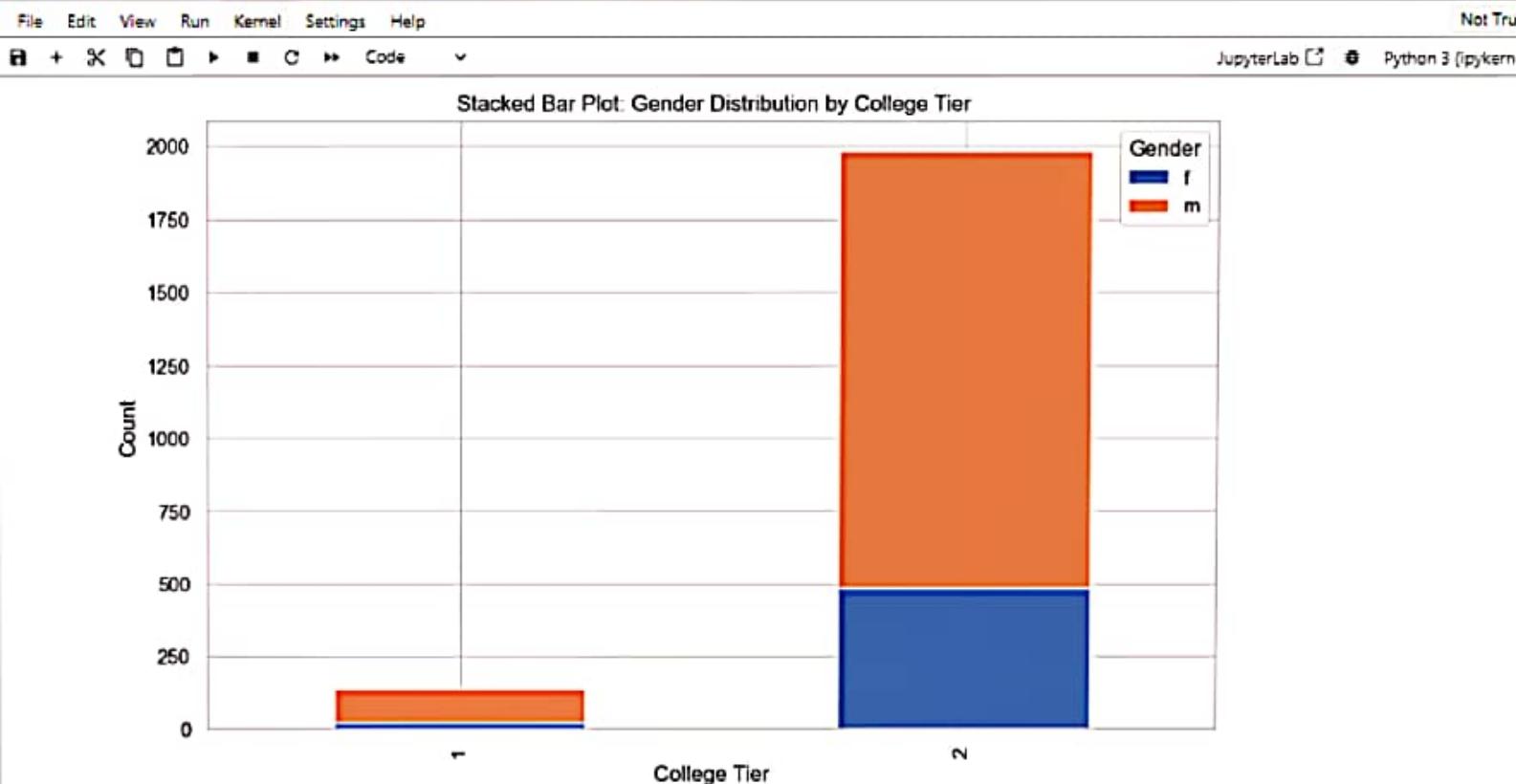
```
[15]: # Bar Plot
import numpy as np
# A bar plot can illustrate the average salary for each designation.
plt.figure(figsize=(12, 6))
sns.barplot(x='Designation', y='Salary', data=df, estimator=np.mean)
plt.title('Bar Plot: Average Salary by Designation')
plt.xticks(rotation=45)
plt.show()

# Observations:
# Identify which designations offer the highest average salaries.
```





jupyter project1_innomatics NEW Last Checkpoint 35 minutes ago



Final Observations

In your final summary, you should encapsulate your findings, including:

Key relationships between numerical variables. Distribution patterns of numerical variables against categorical variables. Insights into how different categories relate to each other, especially concerning salary

Step - 5: Research Questions

Research Question 1: Testing the claim from the Times of India article: The article claims that fresh graduates in fields like Programming Analyst, Software Engineer, Hardware Engineer, and Associate Engineer earn up to 2.5-3 lakhs. We will test this by analyzing the data to check whether the average salary for these job roles aligns with the stated range.

Steps to Test the Claim: Filter the Data for the mentioned designations (Programming Analyst, Software Engineer, Hardware Engineer, and Associate Engineer). Calculate the average salary for these specific roles. Compare the average salary with the claimed range (2.5 to 3 lakhs)

```
[17]: # List of roles mentioned in the claim
roles_to_test = ['programming analyst', 'software engineer', 'hardware engineer', 'associate engineer']

# Filtering data for these roles
df_roles = df[df['Designation'].str.lower().isin(roles_to_test)]

# Calculating the average salary for these roles
```

NOMATICS%20OCTOBER/project1_innomatics%20NEW.ipynb

jupyter project1_innomatics NEW Last Checkpoint: 35 minutes ago Not Trusted

File Edit View Run Kernel Settings Help JupyterLab Python 3 (ipykernel)

```
[37]: # List of roles mentioned in the claim
roles_to_test = ['programming analyst', 'software engineer', 'hardware engineer', 'associate engineer']

# Filtering data for these roles
df_roles = df[df['Designation'].str.lower().isin(roles_to_test)]

# Calculating the average salary for these roles
average_salary = df_roles['Salary'].mean()

# Display the result
print(f"Average Salary for {roles_to_test} roles: {average_salary}")

# This code will give the average salary of graduates with the specified job titles, which can then be compared to the range of 2.5-3 Lakhs.
```

Average Salary for ['programming analyst', 'software engineer', 'hardware engineer', 'associate engineer'] roles: 319816.32653861225

Research Question 2:

Is there a relationship between Gender and Specialization? We want to test if gender influences the choice of specialization, which can be done using a Chi-square test of independence. This will help us understand whether the preference for specialization depends on gender.

Steps to Test the Relationship: Create a contingency table showing the counts of each specialization by gender Perform a Chi-square test to check for independence between the two variables

```
[38]: from scipy.stats import chi2_contingency

# Create a contingency table of Gender and Specialization
contingency_table = pd.crosstab(df['Gender'], df['Specialization'])

# Perform Chi-square test of independence
chi2, p, dof, expected = chi2_contingency(contingency_table)

# Display the test results
print(f"Chi-square statistic: {chi2}")
print(f"P-value: {p}")

# Conclusion based on the p-value
if p < 0.05:
    print("There is a significant relationship between Gender and Specialization (reject null hypothesis).")
else:
    print("No significant relationship between Gender and Specialization (fail to reject null hypothesis).")

# This test will help determine whether there's a statistically significant relationship between gender and specialization.
# If the p-value is less than 0.05, we conclude that the preference for specialization depends on gender.

Chi-square statistic: 78.14474172072177
P-value: 0.00013538755877500643
There is a significant relationship between Gender and Specialization (reject null hypothesis).
```

Step - 6: Conclusion

Based on the undertaken data analysis and hypothesis testing, we can summarize the findings as follows:



File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

Step - 7: Bonus - Interesting Conclusions or Research Questions

Here are some additional research questions and analyses that could uncover more insights from the dataset:

Research Question 1:

Does the level of education (e.g., 10th, 12th, or College GPA) influence salary more significantly?

Objective: To determine which educational stage has the most significant impact on salary.

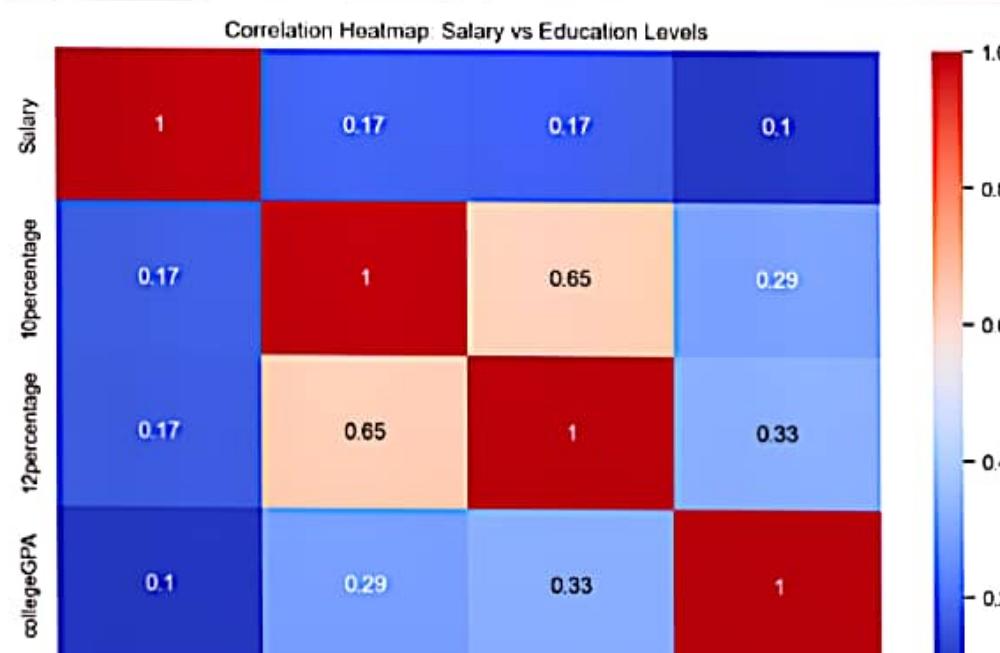
Analysis Approach:

Perform correlation analysis between 10th percentage, 12th percentage, and College GPA with Salary. Visualize the relationships using scatter plots and heatmaps to see which education level correlates more strongly with salary.

```
[39]: plt.figure(figsize=(10, 6))
sns.heatmap(df[['Salary', '10percentage', '12percentage', 'collegeGPA']].corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap: Salary vs Education Levels')
plt.show()
```

Observation:

If college GPA shows the strongest correlation, it would indicate that employers place more importance on a candidate's performance at university rather than high school. # Alternatively, if 12th percentage or 10th percentage shows a stronger correlation, it suggests that early academic performance is a better predictor of future salary.



OMATICS%20OCTOBER/project1_innomatics%20NEW.ipynb



jupyter project1_innomatics NEW Last Checkpoint: 35 minutes ago

File Edit View Run Kernel Settings Help

Not Trusted

A + X D □ ▶ ■ C ** Code

JupyterLab Python 3 (ipykernel)

Research Question 2:

Does the city of employment (JobCity) have a significant impact on salary?

Objective: To determine if the location of a job influences the salary of employees.

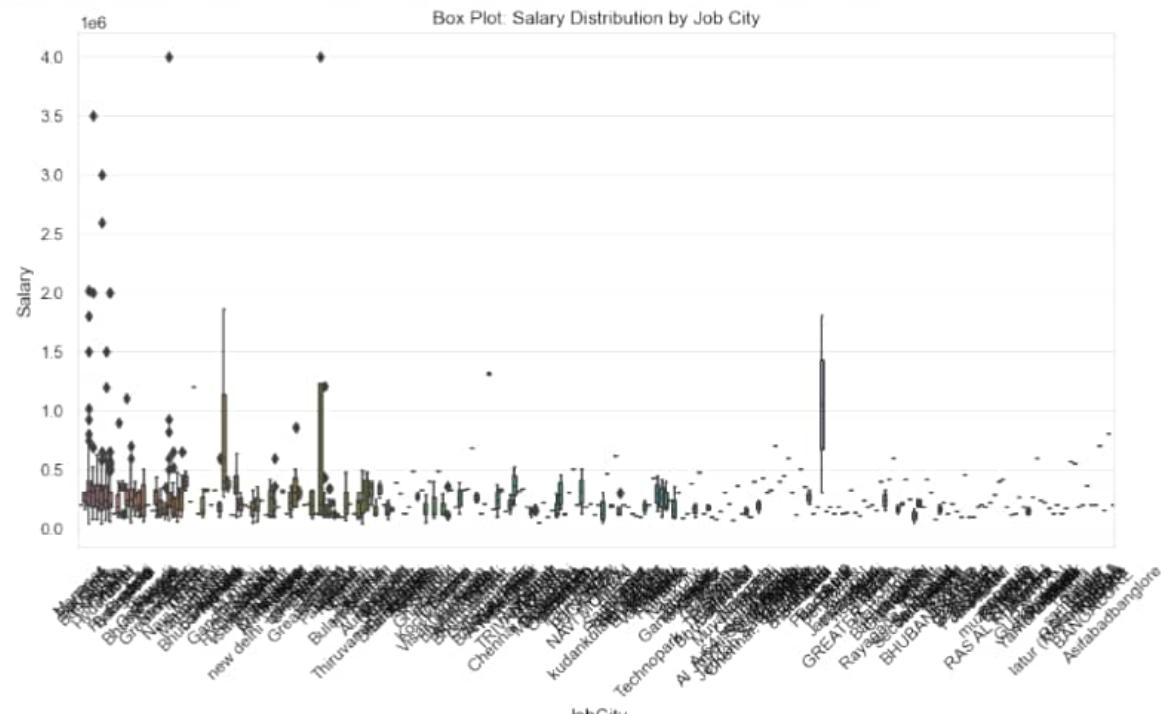
Analysis Approach:

Group the data by JobCity and calculate the average salary for each city. Visualize the results using a bar plot or box plot to compare the salary distributions across different cities.

```
[46]: plt.figure(figsize=(12, 6))
sns.boxplot(x='JobCity', y='Salary', data=df)
plt.title('Box Plot: Salary Distribution by Job City')
plt.xticks(rotation=45)
plt.show()
```

Observation:

Higher average salaries in cities like Bangalore or Mumbai might indicate that tech hubs or metro cities offer higher-paying jobs.
If smaller cities have comparable salaries, it may suggest the spread of opportunities beyond metro areas.



Research Question 3:

Does the number of years of experience (calculated from DOJ and DOL) impact salary?

Objective: To analyze how experience affects salary and if more experience consistently leads to higher pay.

Analysis Approach:

Calculate the experience (in years) from the DOJ and DOL. Use a scatter plot or line plot to visualize the relationship between experience and salary.

```
[41]: df['Experience'] = (pd.to_datetime(df['DOL']) - df['DOJ']).dt.days / 365
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Experience', y='Salary', data=df)
plt.title('Scatter Plot: Salary vs Experience')
plt.xlabel('Experience (years)')
plt.ylabel('Salary')
plt.show()

# Observation:
# A positive correlation between experience and salary would indicate that more experience leads to better salary prospects.
# If the relationship is not linear, it might suggest that other factors, like designation or company, play a stronger role in determining salary.
```

