Cloud_Exploratory_Analysis_Unisys

December 15, 2020

Data Description: The cloud data contains usage and billing summary by certain attributes.

Domain: Cloud instructure

Context: Leveraging client information is paramount for most businesses. In the case of an IT company, attributes of client like the ones mentioned below can be crucial in making business decisions.

Attribute Information * **plan**: Plan of primary description * **usage**: action of using something or the fact of being used * **region**: the beneficiary's business area in the US, India, southeast, northwest * **charges**: Individual costs by client. * **due date**: When the plan expired.

Purpose * Big data and quick anaysis using AL/ML

Learning Outcomes * Exploratory Data Analysis * Practicing statistics using Python * Hypothesis testing

Author: Narender Singh

[9]:
```
# Importing packages
import pandas as pd, numpy as np, scipy.stats as stats, matplotlib.pyplot as␣
 ↪plt, seaborn as sns
import matplotlib.style as style; style.use('fivethirtyeight')
from statsmodels.stats.proportion import proportions_ztest
from statsmodels.formula.api import ols
import statsmodels.api as sm
pd.options.display.max_rows = 4000
from scipy.stats import chi2
```

[2]:
```
# Reading the data as a dataframe
cloud = pd.read_csv('cloudReport.csv')
```

[3]:
```
# first five rows of report file dataframe
cloud.head()
```

[3]:

	Customer	Usage	CPU	Client	Region	Charges	Due Date
0	1	27.900	5	Tata	southwest	16884.92400	1.1.2021
1	2	33.770	3	NASA	northwest	21984.47061	1.1.2022
2	3	33.000	5	Govt	northwest	21984.47061	1.1.2023
3	4	22.705	4	Bank	northwest	21984.47061	1.1.2024
4	5	28.880	0	yes	northwest	3866.85520	1.1.2025

[4]:
```
# Printing out data(shape) in the report
cloud.shape
```

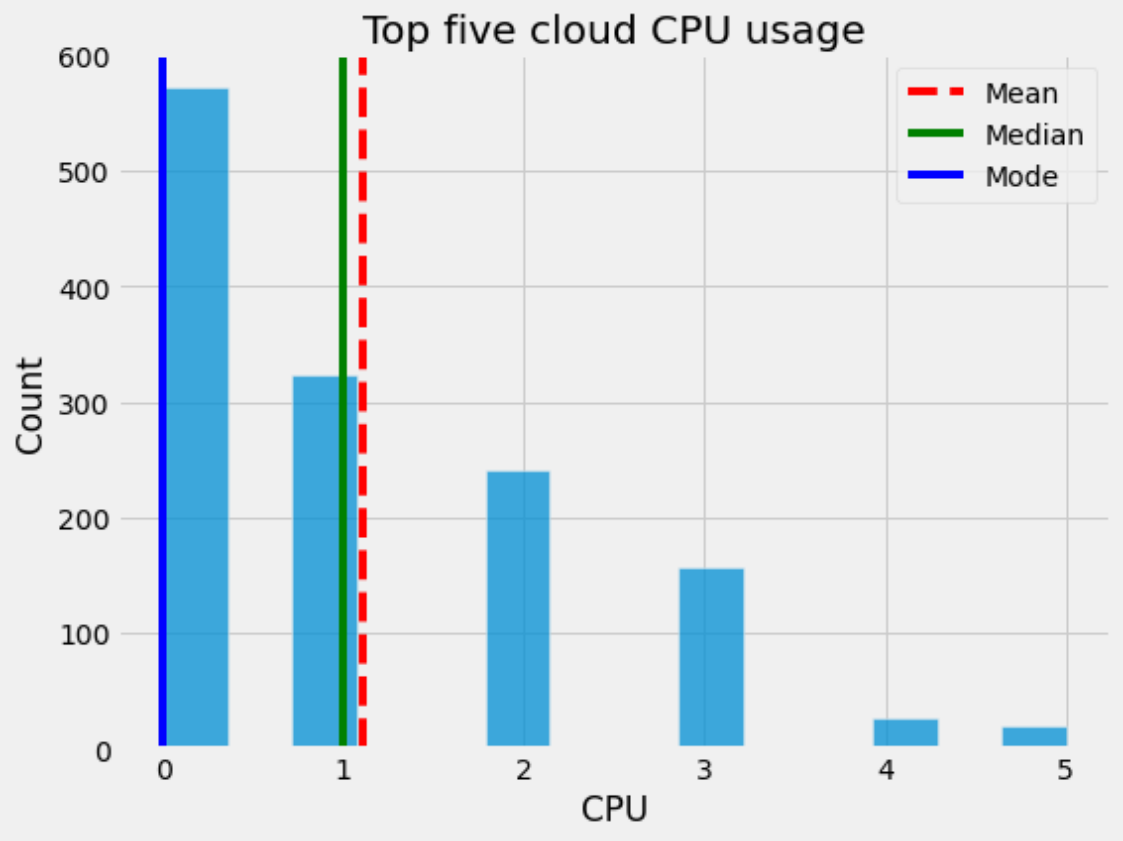
[4]:
```
(1338, 7)
```

[7]:
```
# Cloud CPU usage per five client
cpu_mean = cloud['CPU'].mean()
cpu_median = cloud['CPU'].median()
cpu_mode = cloud['CPU'].mode()

fig, ax_hist = plt.subplots(figsize = (8, 6))
ax_hist = sns.histplot(cloud['CPU'])

ax_hist.axvline(cpu_mean, color = 'r', linestyle = '--', label = 'Mean')
ax_hist.axvline(cpu_median, color = 'g', linestyle = '-', label = 'Median')
ax_hist.axvline(cpu_mode[0], color = 'b', linestyle = '-', label = 'Mode')
ax_hist.set_title('Top five cloud CPU usage')

plt.legend(); plt.show()
```



[8]:
```
# Boxplot for cloud  charges by region.
# Boxplot is a method for graphically depicting groups of numerical data␣
 ↪through their quartiles.
fig = plt.figure(figsize = (12.8, 6))

ax = sns.boxplot(x = 'Region', y = 'Charges', palette = 'afmhot', data = cloud)
ax = sns.stripplot(x = 'Region', y = 'Charges', palette = 'afmhot', data =␣
 ↪cloud,
                   jitter = True, dodge = True, linewidth = 0.5)
ax.set_title('Charges by Region')
```

[8]:
```
Text(0.5, 1.0, 'Charges by Region')
```