

## lab6

December 18, 2024

```
[33]: import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
```

```
[3]: data = pd.read_csv("/Users/yurab/Desktop/Python for university/MMTAD/Lab0/
↳digital_marketing_campaign_dataset.csv")
data
```

```
[3]:
```

	CustomerID	Age	Gender	Income	CampaignChannel	CampaignType	\
0	8000	56	Female	136912	Social Media	Awareness	
1	8001	69	Male	41760	Email	Retention	
2	8002	46	Female	88456	PPC	Awareness	
3	8003	32	Female	44085	PPC	Conversion	
4	8004	60	Female	83964	PPC	Conversion	
...	...	...	...	...	...	...	
7995	15995	21	Male	24849	Email	Awareness	
7996	15996	43	Female	44718	SEO	Retention	
7997	15997	28	Female	125471	Referral	Consideration	
7998	15998	19	Female	107862	PPC	Consideration	
7999	15999	31	Female	93002	Email	Awareness	

	AdSpend	ClickThroughRate	ConversionRate	WebsiteVisits	\
0	6497.870068	0.043919	0.088031	0	
1	3898.668606	0.155725	0.182725	42	
2	1546.429596	0.277490	0.076423	2	
3	539.525936	0.137611	0.088004	47	
4	1678.043573	0.252851	0.109940	0	
...	...	...	...	...	
7995	8518.308575	0.243792	0.116773	23	
7996	1424.613446	0.236740	0.190061	49	
7997	4609.534635	0.056526	0.133826	35	
7998	9476.106354	0.023961	0.138386	49	

7999	7743.627070	0.185670	0.057228	15
------	-------------	----------	----------	----

	PagesPerVisit	TimeOnSite	SocialShares	EmailOpens	EmailClicks	\
0	2.399017	7.396803	19	6	9	
1	2.917138	5.352549	5	2	7	
2	8.223619	13.794901	0	11	2	
3	4.540939	14.688363	89	2	2	
4	2.046847	13.993370	6	6	6	
...	...	...	...	...	...	
7995	9.693602	14.227794	70	13	6	
7996	9.499010	3.501106	52	13	1	
7997	2.853241	14.618323	38	16	0	
7998	1.002964	3.876623	86	1	5	
7999	6.964739	12.763660	2	18	9	

	PreviousPurchases	LoyaltyPoints	AdvertisingPlatform	AdvertisingTool	\
0	4	688	IsConfid	ToolConfid	
1	2	3459	IsConfid	ToolConfid	
2	8	2337	IsConfid	ToolConfid	
3	0	2463	IsConfid	ToolConfid	
4	8	4345	IsConfid	ToolConfid	
...	...	...	...	...	
7995	7	286	IsConfid	ToolConfid	
7996	5	1502	IsConfid	ToolConfid	
7997	3	738	IsConfid	ToolConfid	
7998	7	2709	IsConfid	ToolConfid	
7999	9	341	IsConfid	ToolConfid	

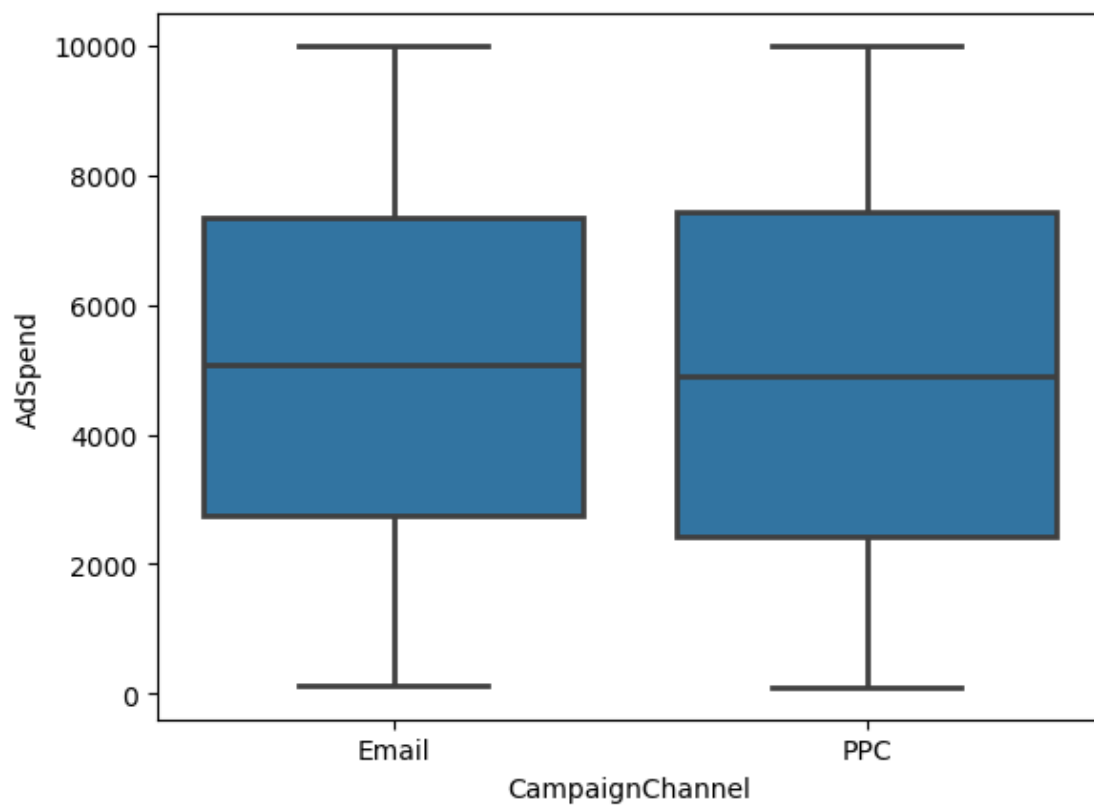
	Conversion
0	1
1	1
2	1
3	1
4	1
...	...
7995	0
7996	0
7997	1
7998	1
7999	0

[8000 rows x 20 columns]

```
[6]: b = pd.concat((data[data["CampaignChannel"] == "Email"],
                    data[data["CampaignChannel"] == "PPC"])))
```

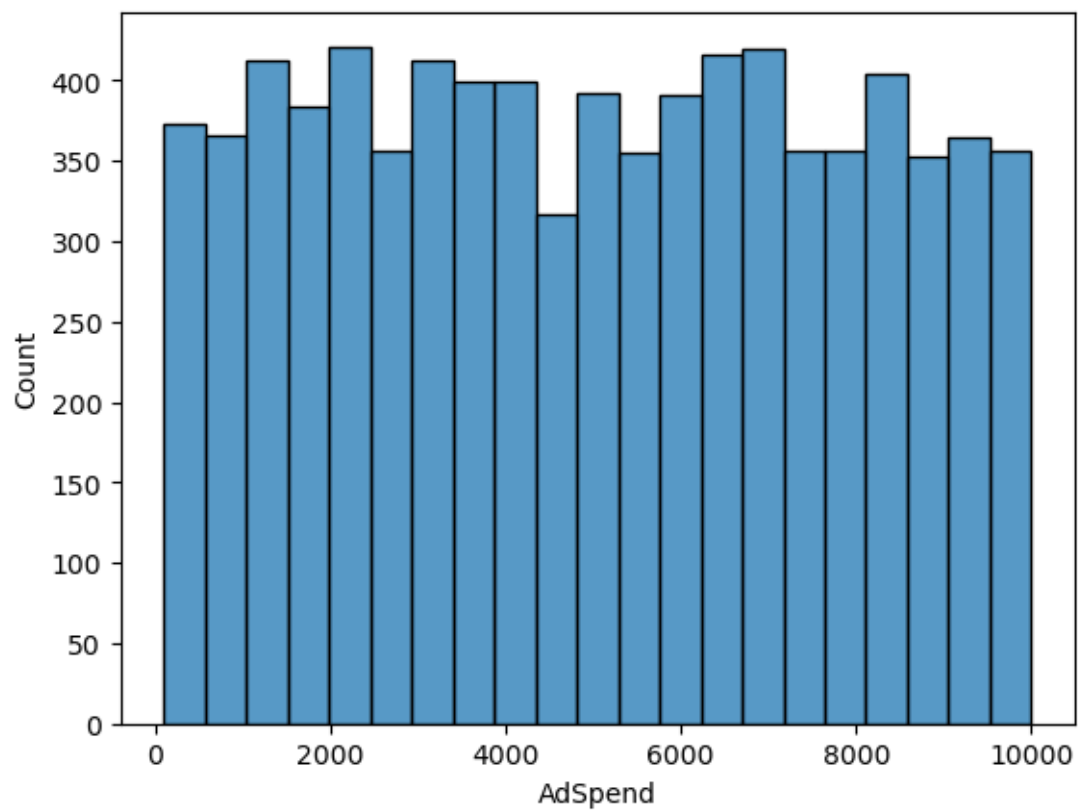
```
[10]: sns.boxplot(y = data["AdSpend"], x = b["CampaignChannel"], linewidth=2)
```

```
[10]: <Axes: xlabel='CampaignChannel', ylabel='AdSpend'>
```



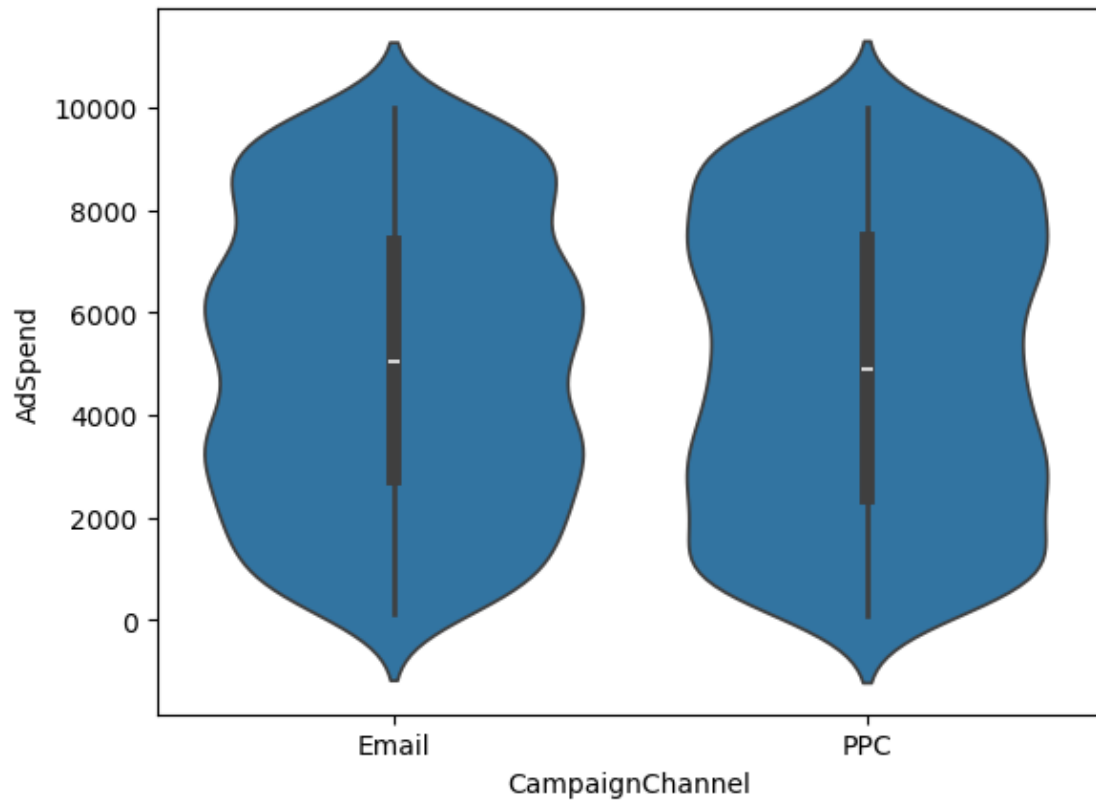
```
[11]: sns.histplot(data['AdSpend'])
```

```
[11]: <Axes: xlabel='AdSpend', ylabel='Count'>
```



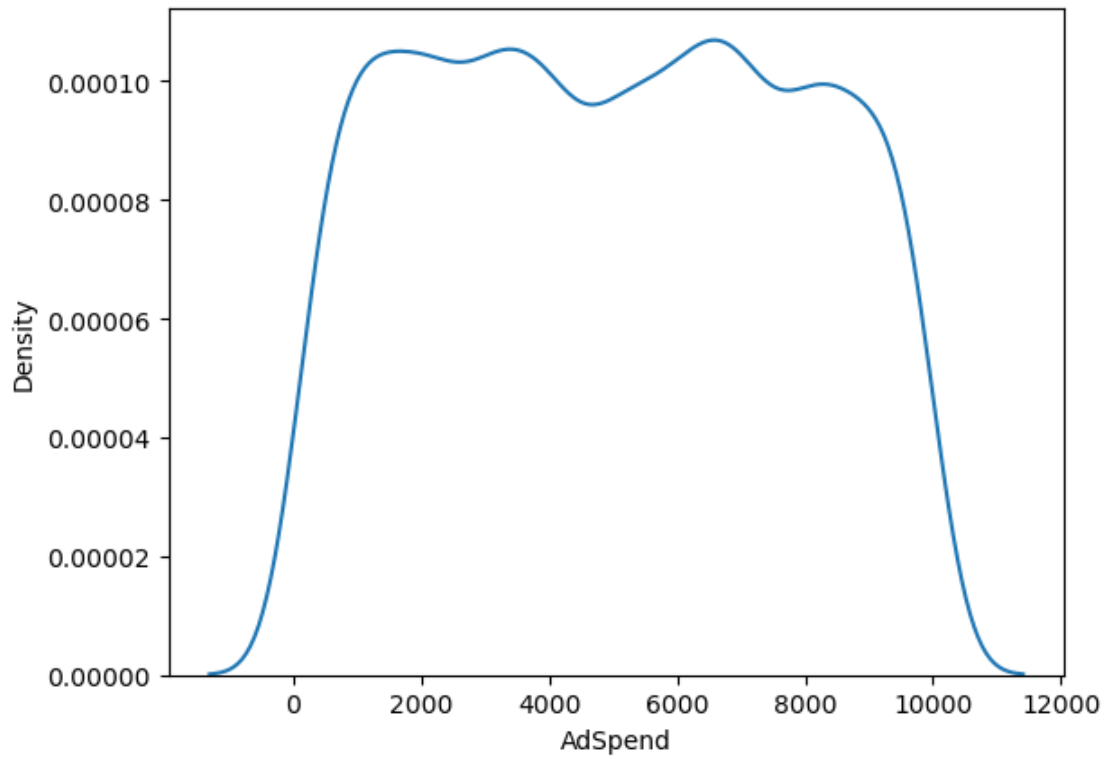
```
[12]: sns.violinplot(x=b["CampaignChannel"], y=data["AdSpend"])
```

```
[12]: <Axes: xlabel='CampaignChannel', ylabel='AdSpend'>
```



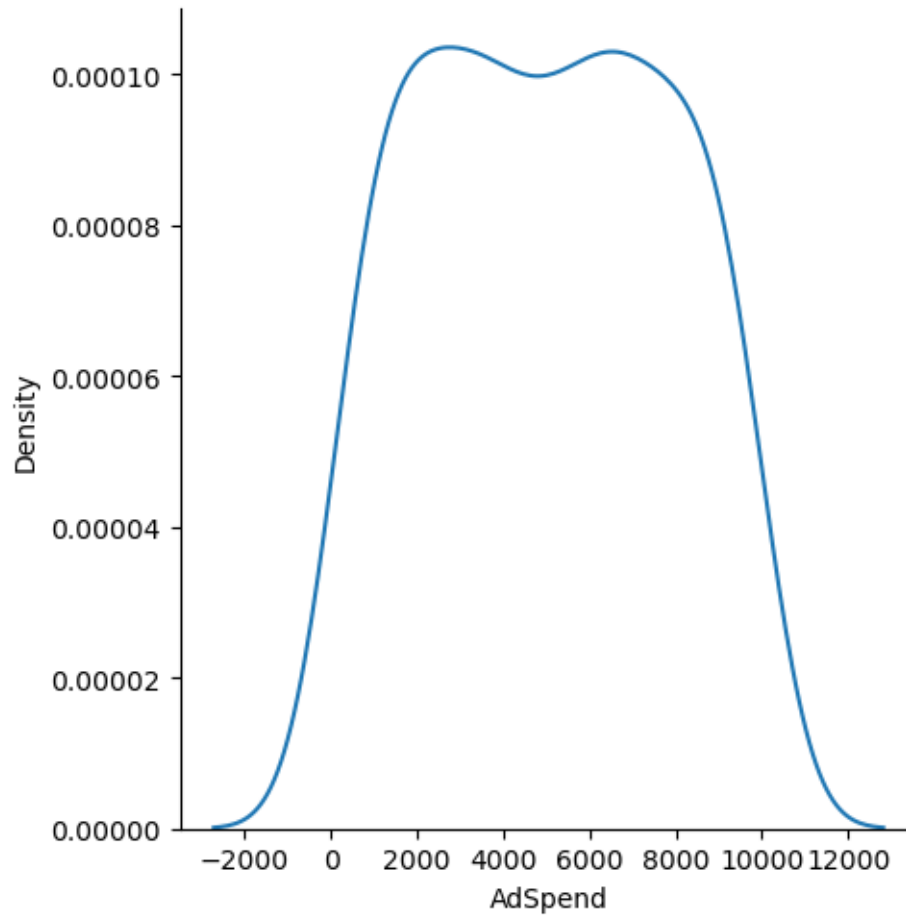
```
[15]: sns.kdeplot(data["AdSpend"])
```

```
[15]: <Axes: xlabel='AdSpend', ylabel='Density'>
```



```
[17]: sns.displot(x=data["AdSpend"], bw_adjust = 2, kind = 'kde')
```

```
[17]: <seaborn.axisgrid.FacetGrid at 0x36f2c09e0>
```



2

```
[18]: data["AdSpend"].corr(data["Age"])
```

```
[18]: -0.0046071461938787285
```

```
[21]: data_without_str = data[["CustomerID", "Age", "Income", "AdSpend",
    ↪ "ClickThroughRate", "ConversionRate", "WebsiteVisits", "PagesPerVisit",
    ↪ "TimeOnSite", "SocialShares", "EmailOpens", "EmailClicks",
    ↪ "PreviousPurchases", "LoyaltyPoints", "Conversion"]]
data_without_str
```

```
[21]:
```

	CustomerID	Age	Income	AdSpend	ClickThroughRate	ConversionRate	\
0	8000	56	136912	6497.870068	0.043919	0.088031	
1	8001	69	41760	3898.668606	0.155725	0.182725	
2	8002	46	88456	1546.429596	0.277490	0.076423	
3	8003	32	44085	539.525936	0.137611	0.088004	

4	8004	60	83964	1678.043573	0.252851	0.109940
...	...	...	...	...	...	...
7995	15995	21	24849	8518.308575	0.243792	0.116773
7996	15996	43	44718	1424.613446	0.236740	0.190061
7997	15997	28	125471	4609.534635	0.056526	0.133826
7998	15998	19	107862	9476.106354	0.023961	0.138386
7999	15999	31	93002	7743.627070	0.185670	0.057228

	WebsiteVisits	PagesPerVisit	TimeOnSite	SocialShares	EmailOpens	\
0	0	2.399017	7.396803	19	6	
1	42	2.917138	5.352549	5	2	
2	2	8.223619	13.794901	0	11	
3	47	4.540939	14.688363	89	2	
4	0	2.046847	13.993370	6	6	
...	...	...	...	...	...	
7995	23	9.693602	14.227794	70	13	
7996	49	9.499010	3.501106	52	13	
7997	35	2.853241	14.618323	38	16	
7998	49	1.002964	3.876623	86	1	
7999	15	6.964739	12.763660	2	18	

	EmailClicks	PreviousPurchases	LoyaltyPoints	Conversion
0	9	4	688	1
1	7	2	3459	1
2	2	8	2337	1
3	2	0	2463	1
4	6	8	4345	1
...	...	...	...	...
7995	6	7	286	0
7996	1	5	1502	0
7997	0	3	738	1
7998	5	7	2709	1
7999	9	9	341	0

[8000 rows x 15 columns]

```
[22]: data_without_str.corr()
```

```
[22]:
```

	CustomerID	Age	Income	AdSpend	ClickThroughRate	\
CustomerID	1.000000	-0.003608	0.002310	0.015931	0.005078	
Age	-0.003608	1.000000	0.008731	-0.004607	0.011542	
Income	0.002310	0.008731	1.000000	0.003668	0.008080	
AdSpend	0.015931	-0.004607	0.003668	1.000000	-0.008296	
ClickThroughRate	0.005078	0.011542	0.008080	-0.008296	1.000000	
ConversionRate	0.010161	0.020027	0.017552	-0.020185	-0.008161	
WebsiteVisits	0.000039	-0.002078	-0.002923	0.007265	-0.023148	
PagesPerVisit	0.000158	-0.009280	0.003868	-0.009256	0.000088	



TimeOnSite	-0.013746	-0.012998	0.019404	-0.004335	-0.008210
SocialShares	-0.011601	-0.009531	-0.006297	-0.021058	-0.013287
EmailOpens	-0.003488	0.015055	-0.001506	0.016514	-0.006076
EmailClicks	-0.003531	0.008820	0.008541	0.001021	-0.010825
PreviousPurchases	-0.012776	-0.007123	-0.011924	0.002378	-0.000387
LoyaltyPoints	-0.023488	0.014049	-0.007181	0.002045	-0.017314
Conversion	-0.180742	0.001606	0.013974	0.124672	0.120012

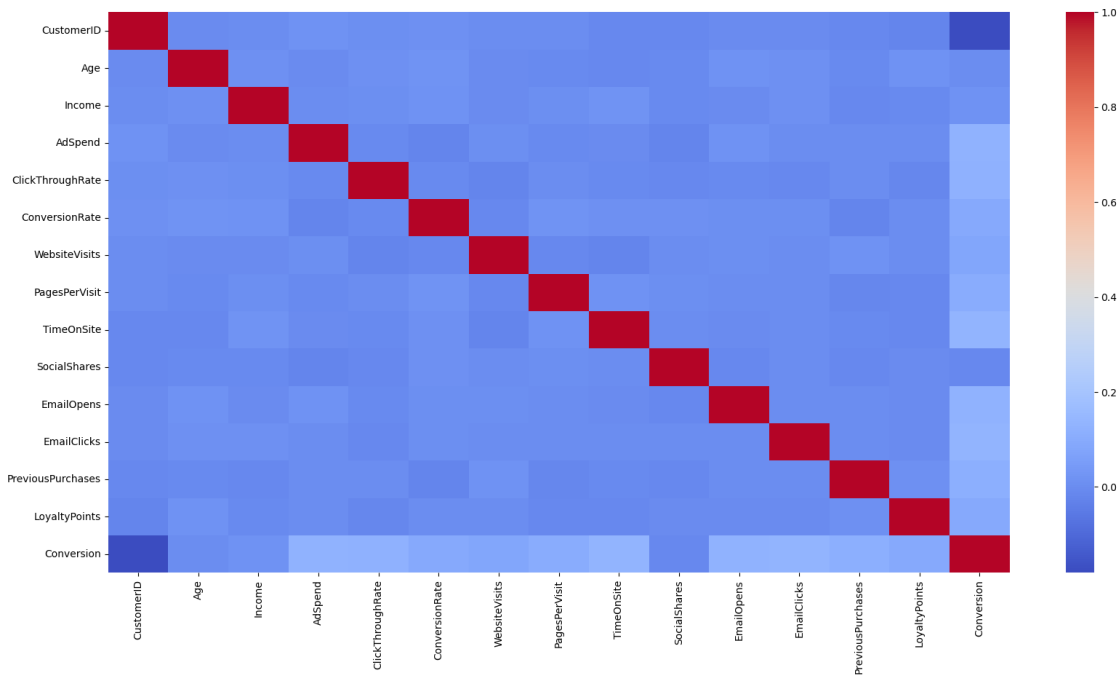
	ConversionRate	WebsiteVisits	PagesPerVisit	TimeOnSite	\
CustomerID	0.010161	0.000039	0.000158	-0.013746	
Age	0.020027	-0.002078	-0.009280	-0.012998	
Income	0.017552	-0.002923	0.003868	0.019404	
AdSpend	-0.020185	0.007265	-0.009256	-0.004335	
ClickThroughRate	-0.008161	-0.023148	0.000088	-0.008210	
ConversionRate	1.000000	-0.012081	0.018789	0.008679	
WebsiteVisits	-0.012081	1.000000	-0.011891	-0.022440	
PagesPerVisit	0.018789	-0.011891	1.000000	0.016455	
TimeOnSite	0.008679	-0.022440	0.016455	1.000000	
SocialShares	0.008713	0.000328	0.007888	0.001869	
EmailOpens	0.006449	0.005865	0.000503	-0.004932	
EmailClicks	0.006716	0.003274	-0.000047	0.002099	
PreviousPurchases	-0.023209	0.013739	-0.016285	-0.006494	
LoyaltyPoints	-0.000498	0.003363	-0.012838	-0.010952	
Conversion	0.093185	0.079339	0.102840	0.129609	

	SocialShares	EmailOpens	EmailClicks	PreviousPurchases	\
CustomerID	-0.011601	-0.003488	-0.003531	-0.012776	
Age	-0.009531	0.015055	0.008820	-0.007123	
Income	-0.006297	-0.001506	0.008541	-0.011924	
AdSpend	-0.021058	0.016514	0.001021	0.002378	
ClickThroughRate	-0.013287	-0.006076	-0.010825	-0.000387	
ConversionRate	0.008713	0.006449	0.006716	-0.023209	
WebsiteVisits	0.000328	0.005865	0.003274	0.013739	
PagesPerVisit	0.007888	0.000503	-0.000047	-0.016285	
TimeOnSite	0.001869	-0.004932	0.002099	-0.006494	
SocialShares	1.000000	-0.012028	0.003214	-0.012635	
EmailOpens	-0.012028	1.000000	0.001410	0.001333	
EmailClicks	0.003214	0.001410	1.000000	0.001284	
PreviousPurchases	-0.012635	0.001333	0.001284	1.000000	
LoyaltyPoints	-0.004575	-0.002839	-0.002837	0.011899	
Conversion	-0.011449	0.124884	0.129521	0.111781	

	LoyaltyPoints	Conversion
CustomerID	-0.023488	-0.180742
Age	0.014049	0.001606
Income	-0.007181	0.013974
AdSpend	0.002045	0.124672

ClickThroughRate	-0.017314	0.120012
ConversionRate	-0.000498	0.093185
WebsiteVisits	0.003363	0.079339
PagesPerVisit	-0.012838	0.102840
TimeOnSite	-0.010952	0.129609
SocialShares	-0.004575	-0.011449
EmailOpens	-0.002839	0.124884
EmailClicks	-0.002837	0.129521
PreviousPurchases	0.011899	0.111781
LoyaltyPoints	1.000000	0.095004
Conversion	0.095004	1.000000

```
[42]: matrix = data_without_str.corr()
plt.figure(figsize=(20,10))
sns.heatmap(matrix, cmap="coolwarm")
plt.show()
```



```
[29]: data[["AdSpend", "Age"]].mode()
```

```
[29]:      AdSpend  Age
0    100.054813  64.0
1    100.668227  NaN
2    100.965939  NaN
```

```

3      103.409243   NaN
4      103.956957   NaN
...
7995   9992.481744   NaN
7996   9996.986533   NaN
7997   9997.002376   NaN
7998   9997.347635   NaN
7999   9997.914781   NaN

```

[8000 rows x 2 columns]

```
[32]: data[["AdSpend", "Age"]].sem()
```

```
[32]: AdSpend    31.730231
      Age       0.166618
      dtype: float64
```

( )

```
[35]: shapiro_test = stats.shapiro(data_without_str)
      print('      - :', shapiro_test)
```

```

      - : ShapiroResult(statistic=0.33243812189315325,
pvalue=1.6237255136214912e-175)

```

```

/opt/anaconda3/lib/python3.12/site-packages/scipy/stats/_axis_nan_policy.py:573:
UserWarning: scipy.stats.shapiro: For N > 5000, computed p-value may not be
accurate. Current N is 120000.

```

```

      res = hypotest_fun_out(*samples, **kwds)
      ,
      ,
      )
      1000

```

```
[40]: bootstrap=pd.DataFrame({"AdSpend": [data.sample(1000, replace=True)["AdSpend"].
      ↪mean() for i in range(0,1000)]})
      bootstrap
```

```
[40]:      AdSpend
0      4956.806302
1      4972.118677
2      4908.787024
3      4888.009462
4      5247.643721
..
995    5054.330041
996    5008.519660
997    4884.037938

```

```
998 5002.595817
999 4910.551925
```

```
[1000 rows x 1 columns]
```

```
[41]: (bootstrap["AdSpend"].quantile(0.025),bootstrap["AdSpend"].quantile(0.975))
```

```
[41]: (4829.01676018463, 5170.764597220722)
```