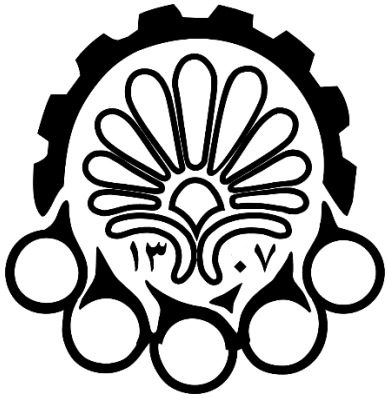


دانشگاه صنعتی امیرکبیر

دانشکده مدیریت علم و فناوری

پروژه اقتصاد سنجی

استاد:

دکتر مهسا جهاندیده

تهیه کننده:

فرید محمدزاده

بهار ۱۴۰۳

مرحله اول:

این دیتاست Voting Outcomes and Campaign Expenditures تحقیقی درباره انتخابات است و متغیر هایی مثل ایالت های مختلف و مناطق مختلف در هر ایالت و اینکه هر نامزد آیا دموکرات است یا نه و چند درصد رای آورده و چقدر برای کمپین انتخاباتی خود هزینه کرده در حالت کلی متغیرهایی از این قبیل. درحالت کلی تمام متغیرهای این تحقیق به عبارت زیر هستند:

- State: postal code of each state that contributed in election
- District: the number of each district in every state
- democA: if the nominate is democrat(=1) or not(=0)
- voteA: the percentage of votes received by candidate A
- expandA: campaign expenditure by candidate A
- expandB: campaign expenditure by candidate B
- prtystrA: the aggregate of most recent votes went to A candidate
- lexpendA: log(expandA)
- lexpendB: log(expandB)
- shareA: the percentage of the campaign expenditure by candidate A out of the sum of the expenditures by candidates A and B

این متغیرها و در حالت کلی این دیتاست میتواند کمک کند که رابطه علّی بین میزان رای آوردن یک نامزد را در انتخابات، با متغیرهایی نظیر مناطق هر ایالت متفاوت، میزان هزینه آنها برای کمپین انتخاباتی خود و میزان سابقه رای های آنها در انتخابات گذشته برسی کند. ممکن است قبل از انجام تحقیق حدس هایی درباره هریک از این متغیر ها و تاثیری که متغیرهای مستقل روی متغیرهای وابسته دارند داشته باشیم. به طور مثال ممکن است تصور کنیم هرچه یک نامزد هزینه بیشتری کرده باشد، شانس دریافت رای بیشتری دارد، یا اینکه میزان هزینه کرد هر نامزد یک متغیر افزاینده کاهشی است. با انجام این تحقیق به سوالاتی مانند این دو سوال میتواند پاسخ داد.

مرحله دوم:

برای انجام هر تحقیقی ابتدا باید متغیرهای مستقلی داشته باشیم تا این متغیر های مستقل متغیر وابسته را توضیح بدهند و اگر نیاز شد که متغیر وابسته پیشبینی شود باید از متغیر های مستقل کمک گرفته بشود. به همین دلیل انتخاب این متغیر های بسیار حائز اهمیت است.

در دیتاست ما متغیرهایی وجود دارند که میتوانند متغیر دیگری را توضیح دهند(متغیرمستقل) و متغیری نیز وجود دارد که توانایی توضیح دادن متغیر های دیگر را ندارد ولیکن از برسی متغیرهای دیگر میتوان رابطه ای میان این دو متغیر پیدا کرد. لیست این متغیر ها به شرح زیر میباشد:

متغیرهای مستقل:

expendA: هزینه‌های کمپین نامزد A ، این متغیر به عنوان یکی از عوامل اصلی تاثیرگذار بر آرای دریافت شده توسط نامزد A مورد بررسی قرار می‌گیرد.

expendB: هزینه‌های کمپین نامزد B ، این متغیر به عنوان یکی از عوامل اصلی تاثیرگذار بر آرای دریافت شده توسط نامزد A مورد بررسی قرار می‌گیرد.

lexpandA: لگاریتم هزینه‌های کمپین نامزد A ، این متغیر برای بررسی تاثیرات غیرخطی هزینه ها استفاده می‌شود.

lexpandB: لگاریتم هزینه‌های کمپین نامزد B ، این متغیر برای بررسی تاثیرات غیرخطی هزینه ها استفاده می‌شود.

shareA: درصد هزینه‌های کمپین نامزد A از مجموع هزینه‌های تبلیغاتی نامزدهای A وB ، این متغیر نشان‌دهنده سهم نسبی هزینه‌های نامزد A نسبت به B است.

تعداد دیگری از متغیرها وجود دارند که به آنها متغیرهای کنترلی یا control variables میگویند و هدف آنها جلوگیری از به وجود آمدن bias است.

state: برای کنترل اثرات منطقه‌ای استفاده می‌شود.

district: برای کنترل اثرات محلی درون ایالت‌ها استفاده می‌شود.

democA: برای کنترل اثرات تفکرات سیاسی استفاده می‌شود.

prtystrA: برای کنترل محبوبیت کلی حزب در انتخابات‌های اخیر استفاده می‌شود.

متغیر وابسته هم عبارت است از:

voteA: این متغیر نشان‌دهنده نتیجه‌ای است که تحت تأثیر متغیرهای دیگر قرار می‌گیرد و هدف اصلی تحلیل آن، این است که ببینیم به طور مثال به ازای هر هزار دلار هزینه ای که روی کمپین تبلیغاتی نامزد A میشود چه میزان رای بیشتری به نامزد A داده میشود و مثال های بیشتری که میتوان برسی کرد.

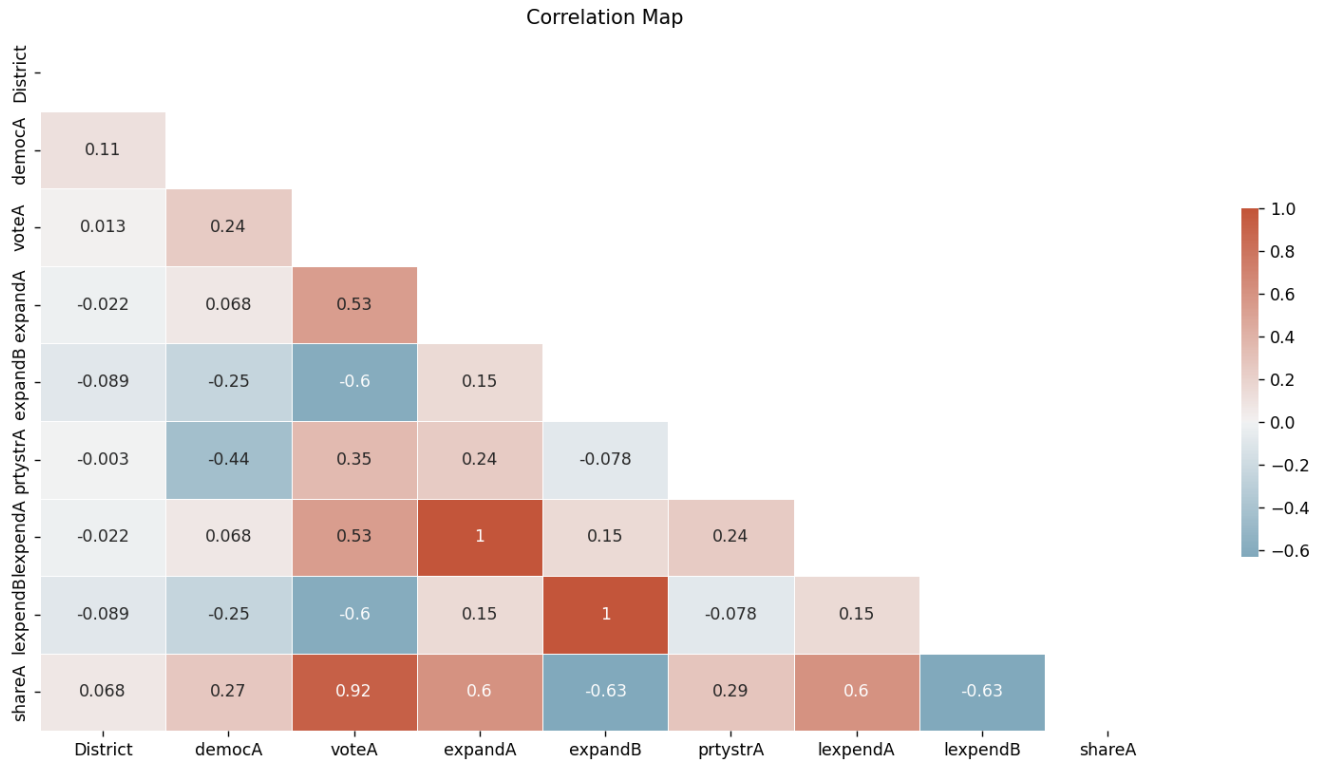
برای تعیین نوع رابطه متغیرهای وابسته و مستقل نیاز است تا تست های آماری روی این نمومه انجام بگیرد تا به لحاظ ریاضی رابطه مستقیم(مثبت یا منفی) آنها مشاهده شود، ولیکن به لحاظ منطقی میتوان انتظار داشت که متغیرهای مستقلی مثل expandA و lexpandA با متغیر وابسته رابطه مستقیم مثبت داشته باشند،

زیرا به صورت منطقی و در حالت کلی هرچه یک نامزد انتخاباتی برای کمپین خود هزینه بیشتری بکند، احتمالا تعداد رای بیشتری میتواند جمع کند و به همین ترتیب پیش بینی میشود که shareA هم رابطه مستقیم مثبتی با متغیر وابسته خواهد داشت.

با استفاده از correlation heatmap میتوانیم رابطه بین متغیر های مستقل و متغیر وابسته و رابطه بین متغیرهای مستقل را ببینینم.

```
df = pd.read_csv("Voting_Outcomes_and_Campaign_Expenditures.csv")
test = df['State']
df.drop(["State"],axis=1,inplace = True)

f, ax = plt.subplots(figsize=(30, 25))
mat = df.corr('spearman')
mask = np.triu(np.ones_like(mat, dtype=bool))
cmap = sns.diverging_palette(230, 20, as_cmap=True)
sns.heatmap(mat, mask=mask, cmap=cmap, vmax=1, center=0,annot = True,
linewidths=.5, cbar_kws={"shrink": .5})
plt.title('Correlation Map')
plt.show()
```

اگر کد بالا را ران کنیم heatmap پایین را میتوانیم خروجی بگیریم.



همانطور که مشخص است و همانطور که انتظار داشتیم رابطه مثبتی بین متغیر های expandA و LexpandA و متغیر وابسته یعنی voteA وجود دارد. و همینطور این رابطه مثبت بین shareA و voteA نیز برقرار است و ما نیز همین انتظار را داشتیم.

رابطه های مثبت و منفی دیگری هم وجود دارد که با توجه به شکل بالا به راحتی میتوان آنها را مشاهده و سپس تحلیل کرد.

در مراحل بعدی به دنبال این هستیم که فرضیات خود را که به لحاظ منطقی رابطه مسقیم دارند برسی کنیم و ببینیم آیا واقعا متغیرهای وابسته میتوانند متغیر مستقل را توضیح دهند و تا چه میزان معنا دار هستند.

مرحله سوم:

در مرحله بعدی مقدمتا داده های categorical خود را (متغیر مستقل state) را به فرمت مناسب در میاوریم تا بتوان آنرا تحلیل کرد. از کد های زیر استفاده میکنیم:

```
df_encoded = pd.get_dummies(df, columns=['State'])
df_encoded = df_encoded.astype(float)
print(df_encoded)
```

داده های متغیر state به صورت {۰ و ۱} در می آیند.

سپس برای بررسی شاخص های پراکندگی و مرکزی از append، describe() استفاده میکنیم تا بتوانیم درک بهتری از داده ها داشته باشیم.

```
print(df_encoded.describe())
```

و خروجی آن به صورت زیر خواهد بود:

	District	democA	voteA	expandA	expandB	prtystrA	...	State_"TX"	State_"UT"	State_"VA"	State_"WA"	State_"WI"	State_"WV"
count	173.000000	173.000000	173.000000	173.000000	173.000000	173.000000	...	173.000000	173.000000	173.000000	173.000000	173.000000	173.000000
mean	8.838150	0.554913	50.502890	310.086705	304.583815	49.757225	...	0.040462	0.011561	0.011561	0.028902	0.023121	0.011561
std	8.768823	0.498418	16.784761	280.969229	306.302993	9.983650	...	0.197613	0.107208	0.107208	0.168017	0.150725	0.107208
min	1.000000	0.000000	16.000000	0.000000	0.000000	22.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	3.000000	0.000000	36.000000	81.000000	60.000000	44.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	6.000000	1.000000	50.000000	242.000000	221.000000	50.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	11.000000	1.000000	65.000000	457.000000	450.000000	56.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
max	42.000000	1.000000	84.000000	1470.000000	1548.000000	71.000000	...	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

در گام بعدی به سراغ تمیز کردن داده ها میرویم تا بعدا بتوانیم آنها را درست تحلیل کنیم و دچار اشتباه نشویم. اولا برای تمیز کردن داده ها، سطر های تکراری دیتاست را حذف میکنیم تا دو سطر مشابه نداشته باشیم:

```
df_encoded = df_encoded.drop_duplicates()
print(df_encoded)
```

که البته از ابتدا سطور تکراری نداشتیم و به همین خاطر دیتاست تغییری نمیکند

سپس داده های پرت را به صورت زیر حذف میکنیم(فرض میکنیم داده های پرت داده هایی هستند که بیش از ۱.۵ دامنه میان چارکی فاصله دارند)

```
df_original = pd.read_csv("Voting_Outcomes_and_Campaign_Expenditures.csv")
df = df_original.copy()

df.drop(["State"], axis=1, inplace=True)
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1

df = df[~((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 * IQR))).any(axis=1)]
df["State"] = df_original.loc[df.index, "State"]

print(df)
```

بعد از اینکه داده های پرت را حذف کردیم تعداد آنها کاهش پیدا میکند

```
     District  democA  voteA  expandA  expandB  prtystrA  lexpendA  lexpendB  shareA State
0           7       1     68   328.30     8.74        41  5.793916  2.167567   97.41  "AL"
1           1       0     62   626.38   402.48        60  6.439952  5.997638   60.88  "AK"
2           2       1     73    99.61     3.07        55  4.601233  1.120048   97.01  "AZ"
3           3       0     69   319.69    26.28        64  5.767352  3.268846   92.40  "AZ"
4           3       0     75   159.22    60.05        66  5.070293  4.095244   72.61  "AR"
..        ...     ...    ...      ...      ...       ...       ...       ...     ...   ...
167         1       0     25    15.22   103.15        49  2.722610  4.636223   12.86  "WV"
168         4       0     39    32.04   152.27        42  3.466954  5.025662   17.38  "WV"
169         3       1     32    22.63   359.80        53  3.119100  5.885551    5.92  "WI"
171         7       0     38   202.59   450.72        46  5.311189  6.110837   31.01  "WI"
172         8       1     30    14.42   227.82        47  2.668685  5.428569    5.95  "WI"

[142 rows x 10 columns]
```

(این مرحله در کد قبل از تغییر متغیر مستقل state به متغیر های ۰ و ۱ انجام شده ولی در توضیحات pdf اینجا آورده شده که مراحل به ترتیب باشد)

در قدم بعدی باید به دنبال داده های گمشده یا اشتباه بگردیم. برای این کار یک تابع تعریف میکنیم که برای ما جستوجو کند و ببیند اساسا آیا داده گمشده یا اشتباهی داریم که با علامت ؟ یا علامت / نمایش داده شده باشد یا خیر.

```
def contains_special_characters(cell):
    if isinstance(cell, str):
        return '?' in cell or '/' in cell
    return False
```

و بعد از اینکه این تابع را در بالای کد نوشتیم، آنرا صدا میزنیم

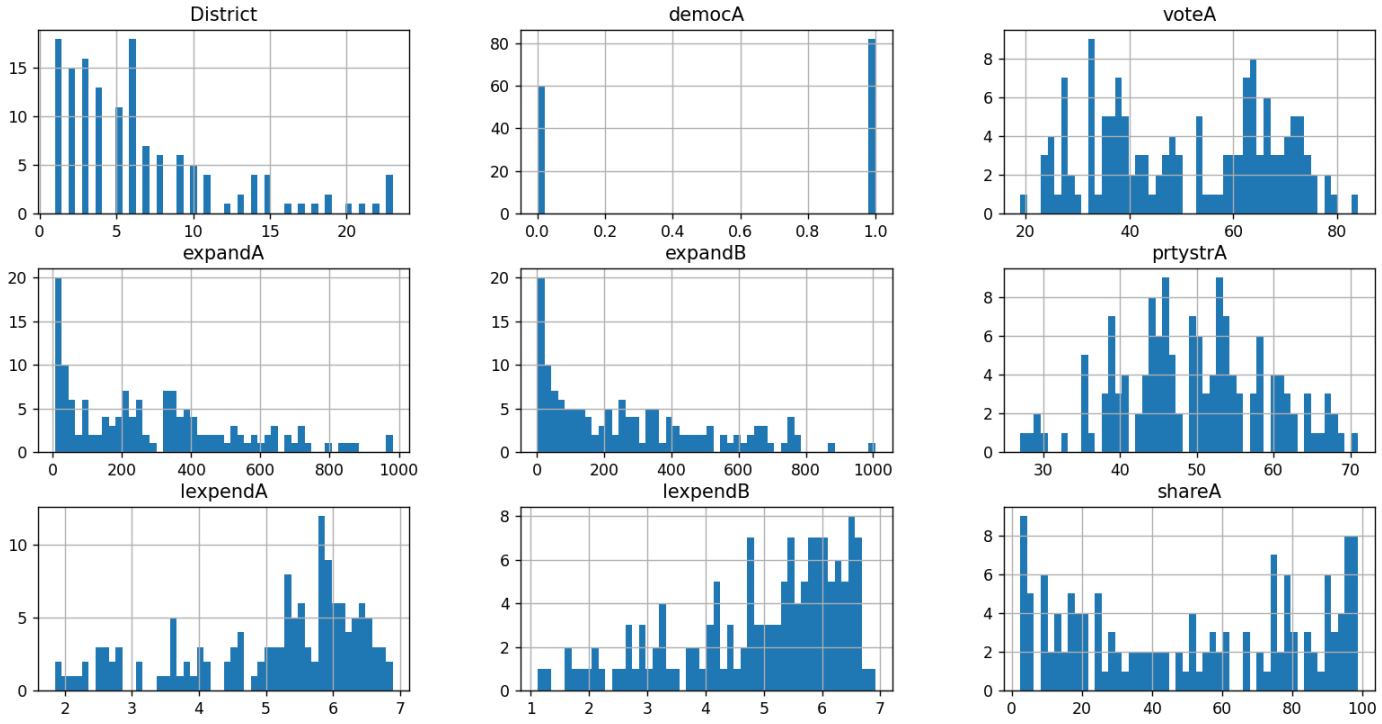
```
contains_characters = df_encoded.applymap(contains_special_characters)
any_special_characters = contains_characters.any().any()

if any_special_characters:
    print("there is some data missing")
else:
    print("all the dataset is complete")


all the dataset is complete
```

مشخص میشود که دیتاست ما کامل است و داده گمشده ای وجود ندارد که نیاز باشد آنرا تخمین بزنیم یا سطر آنرا حذف کنیم.

برای بررسی توزیع داده ها میتوانیم از هیستوگرام استفاده کنیم

```
df.hist(figsize = (35,30), bins = 50 )
plt.show()
```

و خروجی آن به صورت زیر قابل مشاهده است:



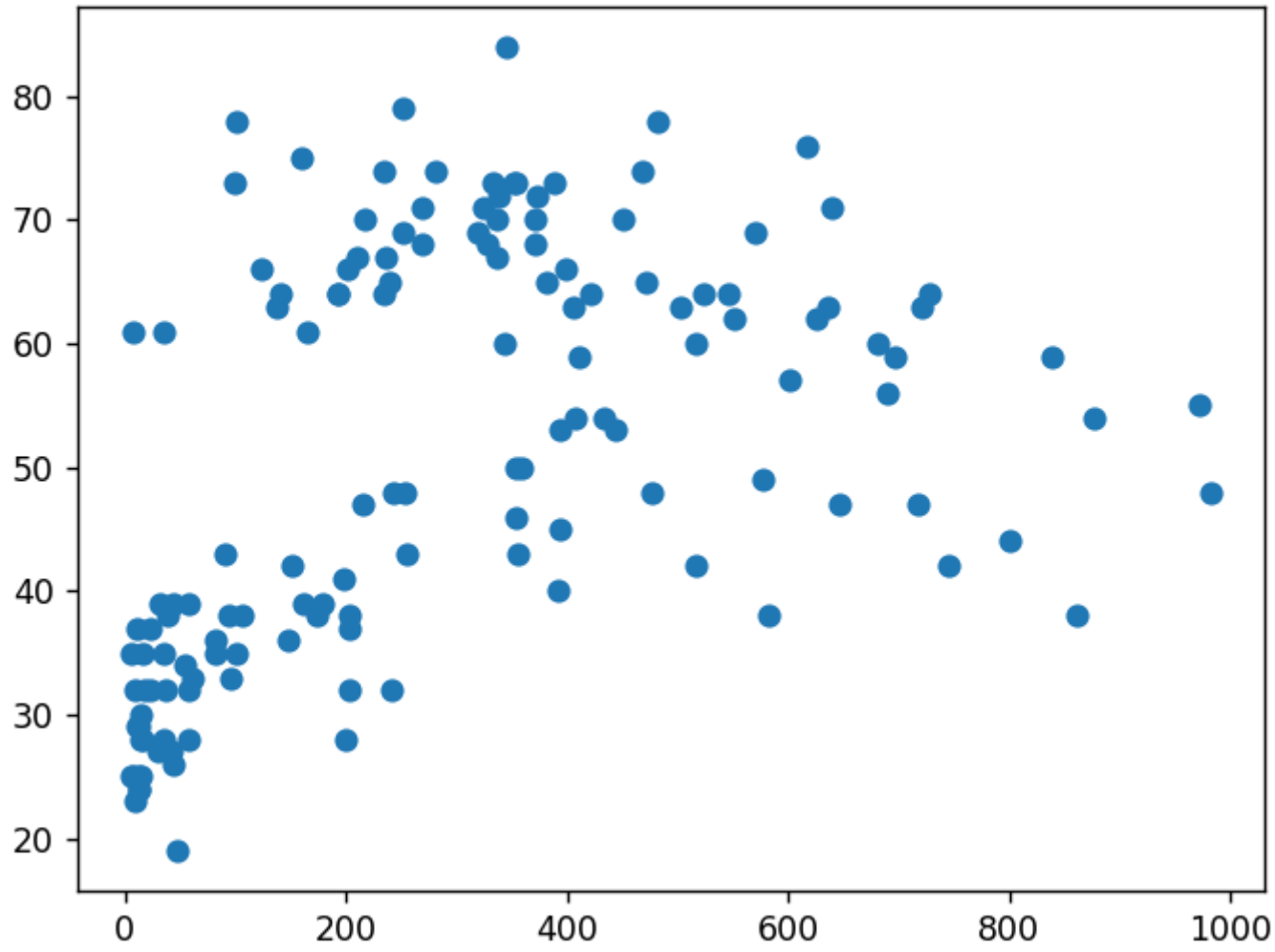
از آنجایی که در دیتاست سطر تکراری و یا داده های گمشده نداشتیم به جای df_encoded از df استفاده میکنیم برای اینکه توزیع ایالت ها برای ما جذابیتی ندارد و اینطور بهتر داده ها visualize میشوند.

و در قدم آخر هم با استفاده از نمودار scatter plot رابطه بین متغیر های مستقل و متغیر وابسته را بررسی میکنیم:

با کمک بین scatter plot رابطه بین هر دو متغیری مورد نظر را میتوان مصور کرد. به طور مثال نمودار زیر رابطه بین expandA و voteA برسی میکند.

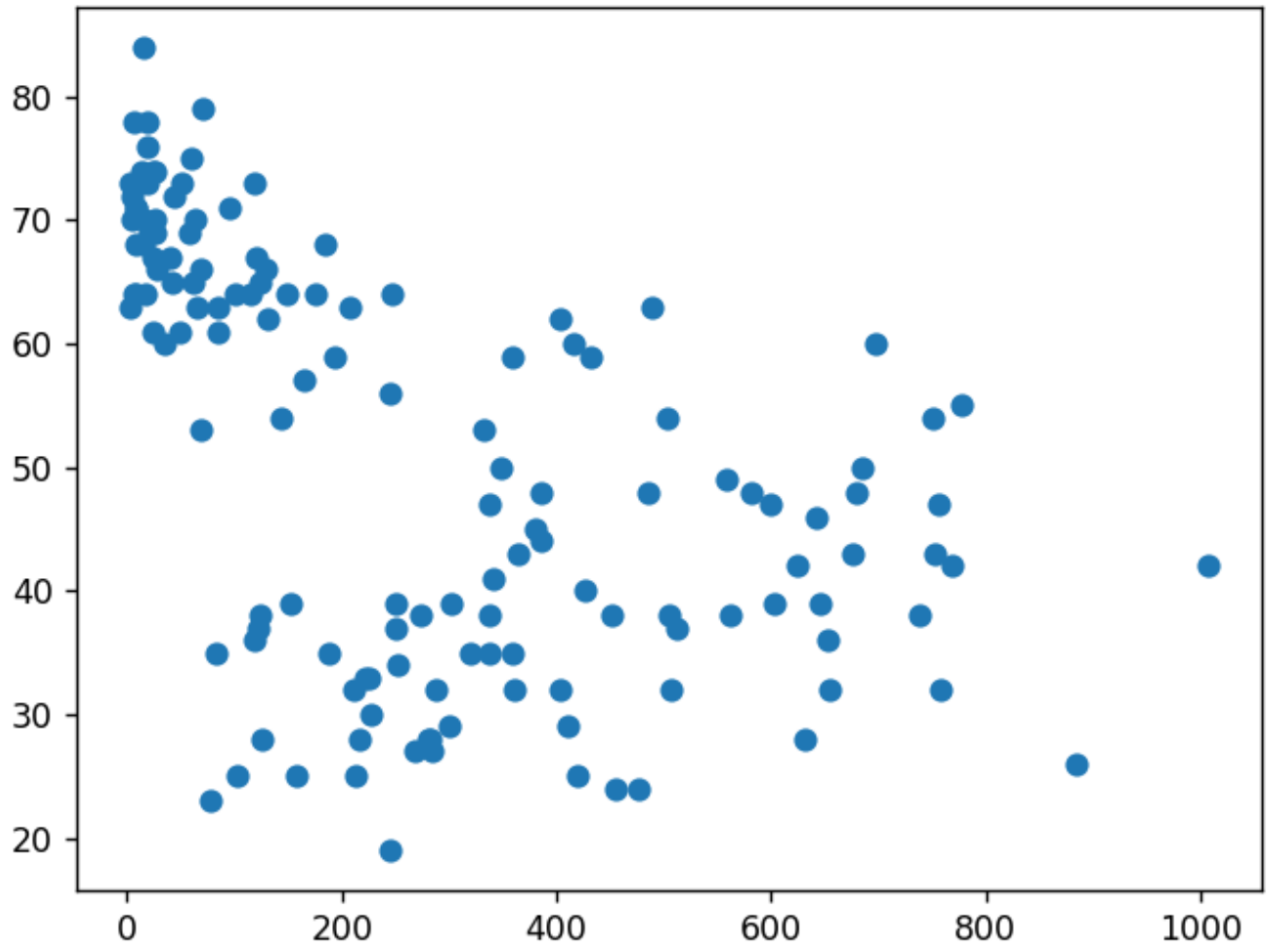
```
plt.scatter(df_encoded['expandA'], df_encoded['voteA'])
plt.show()
```

خروجی آن به صورت زیر خواهد بود:



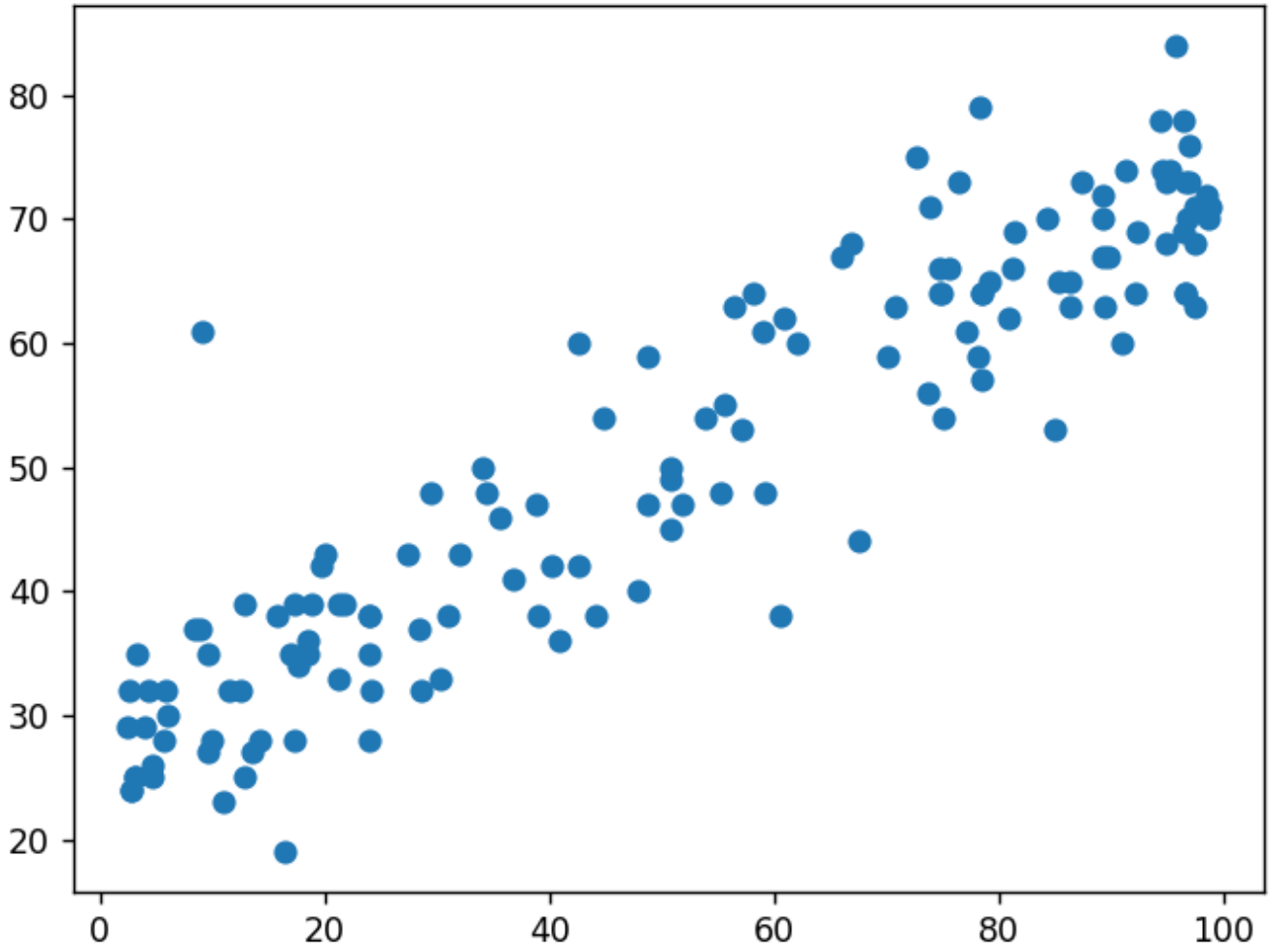
محور افقی نمودار بالا مقدار expandA به هزار دلار است و محور عمودی درصد آرای کاندیدای A است، و همانطور که مشخص است هرچه هزینه های کمپین یک نامزد بیشتر بوده، درصد آرای بالاتری هم کسب کرده.

انتظار میرود اگر نمودار expandB و voteA را رسم کنیم رابطه ای منفی داشته باشند که در شکل زیر مشخص است:



همانطور که از نمودار بالا مشخص است به طور متوسط هرچه expandB بیشتر باشد voteA کمتر خواهد بود و این دقیقا انتظاری است که ما داشتیم.(هرچه هزینه نامزد B بیشتر باشد نامزد A آرای کمتری کسب میکند)

البته این نکته نیز حائز اهمیت است که اگر این نمودار با برداشت ما همخوانی نداشت هم باز جای تعجب نبود، زیرا، ممکن است این bias به وجود بیاید که هزینه های بیشتر در هر ایالت به دلیل وجود جمعیت بیشتر یا موارد از این دست باشد و لزوما صرف داشتن هزینه های بیشتر آرای بیشتری را فراهم نکند/ف به همین سبب خوب است که رابطه بین درصد هزینه نامزد A را (shareA) با voteA نمودار کنیم و نتیجه آنرا ببینیم:



همانطور که به وضوح مشخص است رابطه کاملا مستقیمی بین درصد هزینه های نامزد A و میزان آرای کسب شده توسط این نامزد وجود دارد.

مرحله چهارم:

در این مرحله برای برسی بهتر از تکنیک رگرسیون استفاده میکنیم: (مدل اول)

```
y = df_encoded['voteA']
x = df_encoded.drop(columns=['voteA'])

model = sm.OLS(y,x)
results = model.fit()
print(results.summary())
```

و خروجی زیر را به ما میدهد:

OLS Regression Results

Dep. Variable:	voteA	R-squared:	0.943
Model:	OLS	Adj. R-squared:	0.911
Method:	Least Squares	F-statistic:	29.85
Date:	Fri, 14 Jun 2024	Prob (F-statistic):	2.54e-39
Time:	20:01:22	Log-Likelihood:	-398.16
No. Observations:	142	AIC:	898.3
Df Residuals:	91	BIC:	1049.
Df Model:	50		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
District	-0.0647	0.101	-0.643	0.522	-0.265	0.135
democA	1.1198	1.401	0.799	0.426	-1.664	3.903
expandA	-0.0111	0.004	-2.473	0.015	-0.020	-0.002
expandB	0.0076	0.005	1.539	0.127	-0.002	0.017
prtystrA	0.2161	0.071	3.061	0.003	0.076	0.356
lexpendA	0.1760	1.653	0.107	0.915	-3.107	3.459
lexpendB	0.2138	1.440	0.148	0.882	-2.647	3.074
shareA	0.3422	0.108	3.178	0.002	0.128	0.556
State_"AK"	29.7481	9.425	3.156	0.002	11.026	48.470
State_"AL"	27.2308	8.787	3.099	0.003	9.777	44.685
State_"AR"	32.6172	8.890	3.669	0.000	14.958	50.276
State_"AZ"	26.1694	8.745	2.992	0.004	8.798	43.541
State_"CA"	30.8323	8.626	3.574	0.001	13.698	47.966
State_"CO"	30.4708	8.505	3.583	0.001	13.576	47.366
State_"CT"	30.8211	8.821	3.494	0.001	13.300	48.342
State_"DE"	35.3569	9.795	3.610	0.001	15.900	54.813
State_"FL"	28.4645	7.695	3.699	0.000	13.180	43.749
State_"GA"	28.5984	8.083	3.538	0.001	12.542	44.654
State_"IA"	29.6379	8.908	3.327	0.001	11.943	47.333
State_"ID"	32.4125	9.458	3.427	0.001	13.626	51.199
State_"IL"	24.6426	8.816	2.795	0.006	7.131	42.154
State_"IN"	26.8308	8.250	3.252	0.002	10.443	43.218
State_"KS"	32.6929	8.744	3.739	0.000	15.324	50.062
State_"KY"	33.0791	7.867	4.205	0.000	17.453	48.705
State_"MA"	30.8616	8.905	3.466	0.001	13.174	48.550
State_"MD"	30.0457	9.668	3.108	0.003	10.842	49.250
State_"ME"	28.5569	9.862	2.896	0.005	8.968	48.146
State_"MI"	22.8566	7.628	2.996	0.004	7.704	38.009
State_"MN"	14.3018	7.210	1.984	0.050	-0.021	28.624
State_"MO"	16.0374	7.849	2.043	0.044	0.447	31.628
State_"MT"	17.0289	8.414	2.024	0.046	0.316	33.742
State_"NC"	17.1105	7.357	2.326	0.022	2.497	31.724
State_"NE"	9.5042	9.224	1.030	0.306	-8.818	27.827
State_"NJ"	15.0692	7.678	1.963	0.053	-0.182	30.320
State_"NM"	14.7749	8.079	1.829	0.071	-1.273	30.822
State_"NV"	11.6454	8.989	1.296	0.198	-6.210	29.501
State_"NY"	15.5884	8.033	1.940	0.055	-0.369	31.546
State_"OH"	8.7031	7.326	1.188	0.238	-5.849	23.255
State_"OK"	18.5498	8.223	2.256	0.026	2.216	34.883
State_"OR"	9.0057	8.616	1.045	0.299	-8.108	26.120
State_"PA"	15.9320	7.674	2.076	0.041	0.688	31.176
State_"RI"	10.8223	9.805	1.104	0.273	-8.653	30.298
State_"SC"	13.4703	8.495	1.586	0.116	-3.405	30.345
State_"SD"	3.5212	8.867	0.397	0.692	-14.093	21.135
State_"TN"	5.2292	9.262	0.565	0.574	-13.168	23.626
State_"TX"	17.1853	7.688	2.235	0.028	1.914	32.456
State_"UT"	15.7107	8.204	1.915	0.059	-0.585	32.006
State_"VA"	9.4526	8.073	1.171	0.245	-6.584	25.489
State_"WA"	14.7229	7.630	1.930	0.057	-0.432	29.878
State_"WI"	13.9403	7.985	1.746	0.084	-1.921	29.802
State_"WV"	14.8739	7.495	1.984	0.050	-0.014	29.762

Omnibus:	8.523	Durbin-Watson:	2.408
Prob(Omnibus):	0.014	Jarque-Bera (JB):	17.599
Skew:	-0.005	Prob(JB):	0.000151
Kurtosis:	4.725	Cond. No.	5.78e+04

در مدل اول مقدار R-squared برابر ۰.۹۴۳ است، به این معنی که ۹۴.۳ درصد تغییرات در رای های کسب شده توسط نامزد A توسط متغیر های وابسته توضیح داده شدند. مقدار F-statistic نشان میدهد که مجموع تمام متغیر های مستقل کنار هم معنادار است و میتواند متغیر وابسته را توضیح دهد. مقدار P_value برای بعضی از متغیرها کمتر از ۰.۰۵ است که نشان میدهد وجود آنها معنادار است و نمیتوان فرض صفر را در کرد.

برای مدل بعدی متغیر های دیگری را در نظر میگیریم تا بتوانیم بهتر استنتاج کنیم: (مدل دوم)

```
y = df_encoded['voteA']
x = df_encoded[['lexpendA','prtystrA','democA']]

model = sm.OLS(y,x)
results = model.fit()
print(results.summary())
```

و خروجی زیر را میدهد:

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                   voteA   R-squared (uncentered):                   0.951
Model:                             OLS   Adj. R-squared (uncentered):              0.950
Method:                  Least Squares   F-statistic:                              892.4
Date:                 Fri, 14 Jun 2024   Prob (F-statistic):                    1.42e-90
Time:                         20:26:47   Log-Likelihood:                         -553.00
No. Observations:                  142   AIC:                                      1112.
Df Residuals:                      139   BIC:                                      1121.
Df Model:                            3
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
lexpendA       6.2484      0.769      8.120      0.000       4.727       7.770
prtystrA       0.2937      0.073      4.045      0.000       0.150       0.437
democA         8.3869      2.058      4.075      0.000       4.318      12.456
==============================================================================
Omnibus:                       11.730   Durbin-Watson:                   0.875
Prob(Omnibus):                  0.003   Jarque-Bera (JB):                4.938
Skew:                           0.180   Prob(JB):                       0.0847
Kurtosis:                       2.160   Cond. No.                         105.
==============================================================================
```

همانطور که مشخص است رابطه کاملا مستقیم و مثبتی بین متغیرهای مستقل و متغیر وابسته وجود دارد. مقدار R-squared برابر ۰.۹۴۳ است، به این معنی که ۹۴.۳ درصد تغییرات در رای های کسب شده توسط نامزد A توسط متغیر های وابسته توضیح داده شدند. مقدار F-statistic نشان میدهد که مجموع تمام متغیر های مستقل کنار هم معنادار است و میتواند متغیر وابسته را توضیح دهد. مقدار P_value برای هر سه متغیر مستقل نزدیک به ۰ است که نشان میدهد این متغیرها معنادار هستند.

مرحله پنجم:

همانطور که از ابتدا انتظار داشتیم رابطه مستقیمی بین متغیر های مستقل و وابسته در مدل دوم وجود دارد میتوان نشان داد که در صورت ثابت نگه داشتن prtystrA و democA به طور متوسط به ازای هر یک درصد افزایش در هزینه های کمپین، نامزد A ۶.۲۴ درصد رای بیشتری کسب کرده و از طرفی در صورت ثابت بودن دو متغیر prtystrA و lexpendA ایالت های دموکرات به طور متوسط ۸.۳۸ درصد رای بیشتری به نامزد A دادند.

از آنجایی که مقدار t-test همه متغیرهای مستقل مقداری بیشتر از ۱.۶۵ است تمامی آنها معنادار هستند و دلیلی برای رد فرض صفر(coef = 0) وجود ندارد.

میتوانیم علاوه بر آزمون t از مقدار P_value متغیرها هم متوجه شویم که تمامی آنها در سطح اطمینان ۹۵ درصد و حتی ۹۹ درصد معنادار هستند.

از قسمت راست جدول هم میتوان تخمینی برای بازه ضرایب متغیرهای مستقل(در سطح اطمینان ۹۵ درصد) مشاهده کرد.

مقدار adj .R-squared هم ۹۵ درصد است.(برای هر متغیر مستقلی که به مدل اضافه میکنیم مقداری جریمه میکند)

میتوان از این دیتاست و این تحقیق نتیجه گرفت که رابطه مستقیمی بین هزینه های تبلیغاتی نامزد های انتخاباتی و میزان آرایی که جمع میکنند وجود دارد. نکته مهمی که وجود دارد این است که این تاثیر به صورت رابطه مثبت کاهشی است، برای نشان دادن این ادعای از رگرسیون زیر استفاده میکنیم:

OLS Regression Results

Dep. Variable:	voteA	R-squared (uncentered):	0.946
Model:	OLS	Adj. R-squared (uncentered):	0.945
Method:	Least Squares	F-statistic:	1219.
Date:	Fri, 14 Jun 2024	Prob (F-statistic):	2.69e-89
Time:	21:15:56	Log-Likelihood:	-559.77
No. Observations:	142	AIC:	1124.
Df Residuals:	140	BIC:	1129.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
lexpendA	11.2195	0.429	26.134	0.000	10.371	12.068
expandA	-0.0213	0.006	-3.555	0.001	-0.033	-0.009

Omnibus:	2.632	Durbin-Watson:	0.949
Prob(Omnibus):	0.268	Jarque-Bera (JB):	2.079
Skew:	0.152	Prob(JB):	0.354
Kurtosis:	2.491	Cond. No.	154.

همانطور که مشخص است این رابطه مثبت در ابتدا افزایش میابد ولیکن رفته رفته مقدار این افزایش کمتر میشود. مانند نمودار زیر: