

美国选民总统喜好分析

Base on Pandas

李昀哲 20123101

Thursday, Oct 18, 2022

01

Part One

Where to start

- Background
- Datasets
- Charts Merge

Where to start?

- Background

Individuals

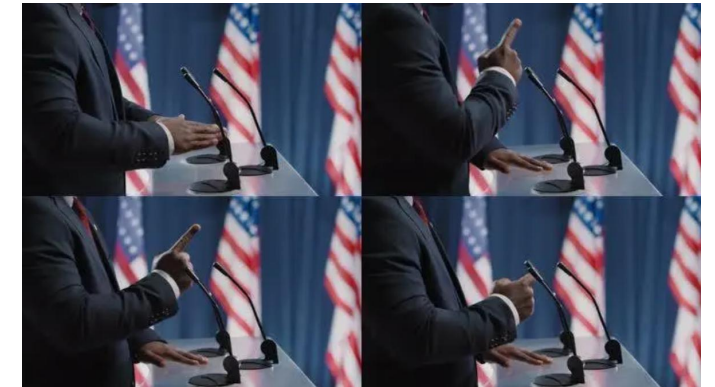


Contribute to

Candidate Committees



President Candidate



Favor & Contribution

Where to start?

- Datasets

File_name	Description
ccl22	候选人委员会相关信息。包含如候选人身份证信息、候选人选举年份、委员会类型等等。
weball22	候选人相关信息。包含候选人身份证姓名、身份证信息、候选人委员会、财务情况、政党等等。
<u>itcont_2022_20220629_20220817</u>	个人捐款信息。包括个人姓名、所属职业、所在州、捐款数量等等。

```
pandas.read_csv(file_path, sep, names)
```

- file_path为文件路径;
- sep为分隔符, default为',';
- names为列名列表;

Where to start?

- Charts Merge

File_name	Description
ccl22	无候选人姓名 候选人委员会相关信息。包含如候选人身份证信息、候选人选举年份、委员会类型等等
weball22	候选人相关信息。包含候选人身份证姓名、身份证信息、候选人委员会、财务情况、政党等等
itcont_2022_20220629_20220817	个人捐款信息。包括个人姓名、所属职业、所在州、捐款数量等等

	CMTE_ID	CAND_ID	CAND_NAME	CAND_PTY_AFFILIATION	(诸多委员会信息)
0	C00697789	H0AL01055	CARL, JERRY LEE, JR	REP	
1	C00701557	H0AL01063	LAMBERT, DOUGLAS WESTLEY III	REP	
2	C00708867	H0AL01097	AVERHART, JAMES	DEM	
3	C00462143	H0AL02087	ROBY, MARTHA	REP	
4	C00493783	H0AL02087	ROBY, MARTHA	REP	

	CAND_NAME	NAME	STATE	EMPLOYER	OCCUPATION	TRANSACTION_AMT	TRANSACTION_DT	CAND_PTY_AFFILIATION
0	CARL, JERRY LEE, JR	DUNN, TIM	TX	CROWN QUEST OPERATING	CFO	2900	6302022	REP
1	CARL, JERRY LEE, JR	HOWELLS, DIANNE M.	TX	MILTOPE	PRESIDENT	500	6302022	REP
2	CARL, JERRY LEE, JR	MCBRIDE, TIMOTHY J.	DC	ST ENGINEERING	PRESIDENT	500	6302022	REP
3	CARL, JERRY LEE, JR	AGUA CALIENTE BAND OF CAHUILLA INDIANS	CA	NaN	NaN	2900	6302022	REP
4	CARL, JERRY LEE, JR	PECHANGA BAND OF INDIANS	CA	NaN	NaN	2900	6302022	REP
5	CARL, JERRY LEE, JR	POARCH BAND OF CREEK INDIANS	AL	NaN	NaN	2900	6302022	REP
6	HARVEY-HALL, PHYLLIS	BISSOO, MIRANDA	AL	NaN	NaN	500	6302022	DEM
7	SEWELL, TERRI A.	ELAM, GARY	AL	NOT EMPLOYED	NOT EMPLOYED	100	6292022	DEM
8	SEWELL, TERRI A.	BLOOM, STAR	AL	NOT EMPLOYED	NOT EMPLOYED	250	6292022	DEM
9	SEWELL, TERRI A.	MCCOY, MOYER	DC	PMI GLOBAL SERVICES	MANAGER EXTERNAL AFFAIRS	350	6292022	DEM
10	SEWELL, TERRI A.	SULLIVAN, CONNIE	CO	NHIA	PHARMACIST	500	6292022	DEM
11	SEWELL, TERRI A.	WOODS, KATHRYN	CA	GENETECH	PRINCIPAL PROGRAMMER ANALYST	625	6302022	DEM

02

Part Two

Data Process & Analysis

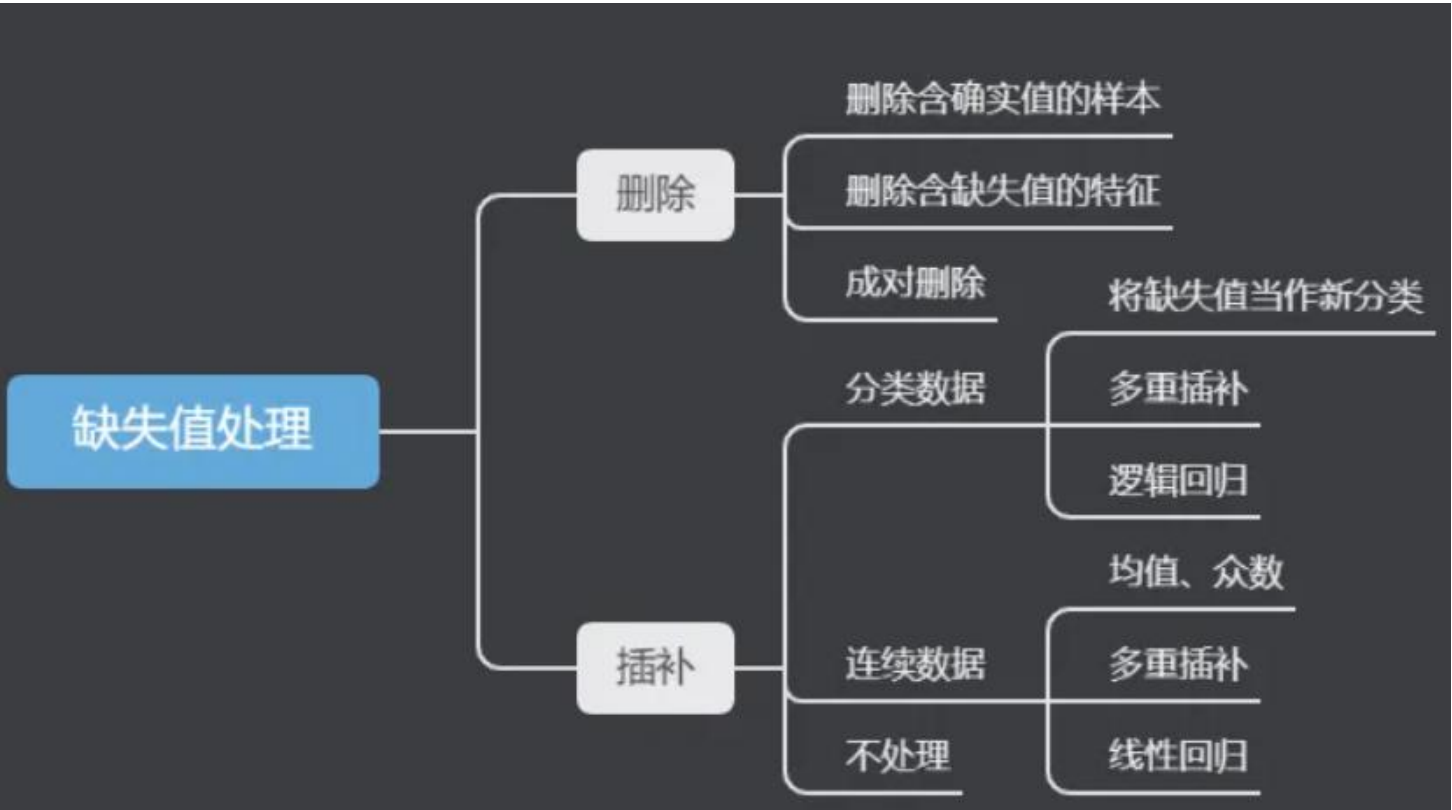
- 填充空值
- 分析数据

Data Process & Analysis

- 空值填充

`dataframe.info()`

```
Data columns (total 8 columns):
#      Column      Non-Null
---  -
0     CAND_NAME    326055 no
1     NAME         326055 no
2     STATE        325980 no
3     EMPLOYER     308358 no
4     OCCUPATION   313541 no
5     TRANSACTION_AMT 326055 no
6     TRANSACTION_DT 326055 no
7     CAND_PTY_AFFILIATION 326055 no
dtypes: int64(2), object(6)
```



Data Process & Analysis

- 空值填充

```
cand_cmte_indi['STATE'].fillna('NOT PROVIDED',inplace=True)  
cand_cmte_indi['EMPLOYER'].fillna('NOT PROVIDED',inplace=True)  
cand_cmte_indi['OCCUPATION'].fillna('NOT PROVIDED',inplace=True)
```

Data Process & Analysis

- 数据分析

```
DataFrame.groupby('Label_Name').sum().sort_values("Label_Name")
```

党派情况

CAND_PTY_AFFILIATION	TRANSACTION_AMT
DEM	97503585
REP	29409549
DFL	500940
IND	297825
OTH	291199
UN	72957
LIB	25543
NPA	19793
UNK	14202
GRE	10850

候选人情况

CAND_NAME	TRANSACTION_AMT
MERCER, LEE	64000000
SULLIVAN, DAN	5092362
DEMINGS, VAL	2705822
HASSAN, MARGARET WOOD	2481407
KELLY, MARK	1790196
RUBIO, MARCO	1507452
PELOSI, NANCY	1135776
RYAN, PATRICK	980479
WEST, MATTHEW	886669
EVANS, JAKE	814331



Data Process & Analysis

- 数据分析

	CAND_NAME	NAME	STATE	EMPLOYER	OCCUPATION	TRANSACTION_AMT	TRANSACTION_DT	CAND_PTY_AFFILIATION
96230	MERCER, LEE	MERCER, LEE L MR. JR.	TX	S.GOVERNMENT/CONGRESS U. TOO	INTELLIGENCE OFFICER	64000000	2022630	DEM

Total number: 326,055

SULLIVAN, DAN	71427
DEMINGS, VAL	43867
RUBIO, MARCO	21961
KELLY, MARK	21916
HASSAN, MARGARET WOOD	14831
OCASIO-CORTEZ, ALEXANDRIA	8669
SANDERS, BERNARD	7084
WARNOCK, RAPHAEL	6810
JOHNSON, RON HAROLD MR.	5682
STEFANIK, ELISE M.	4116

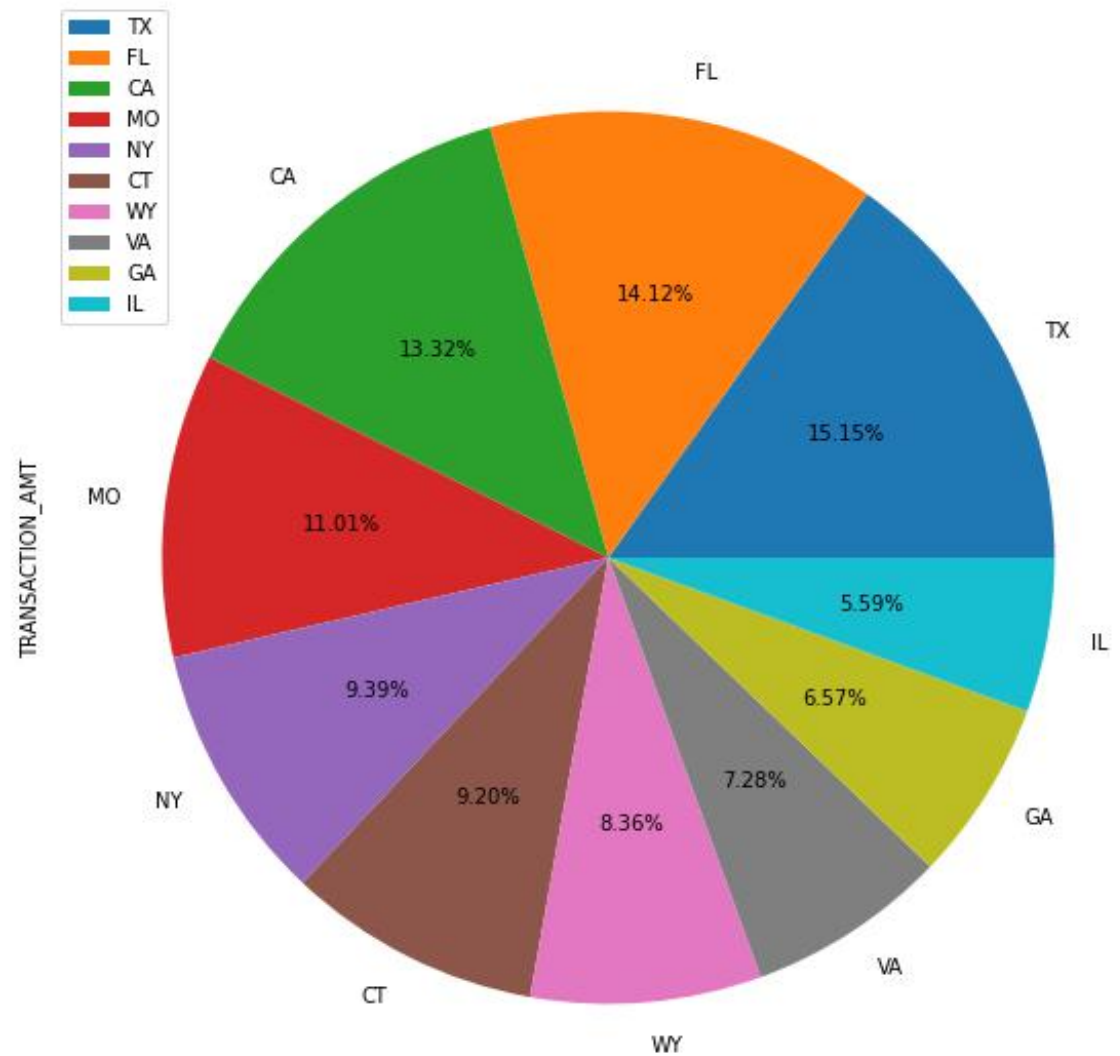
03

Part Theeee

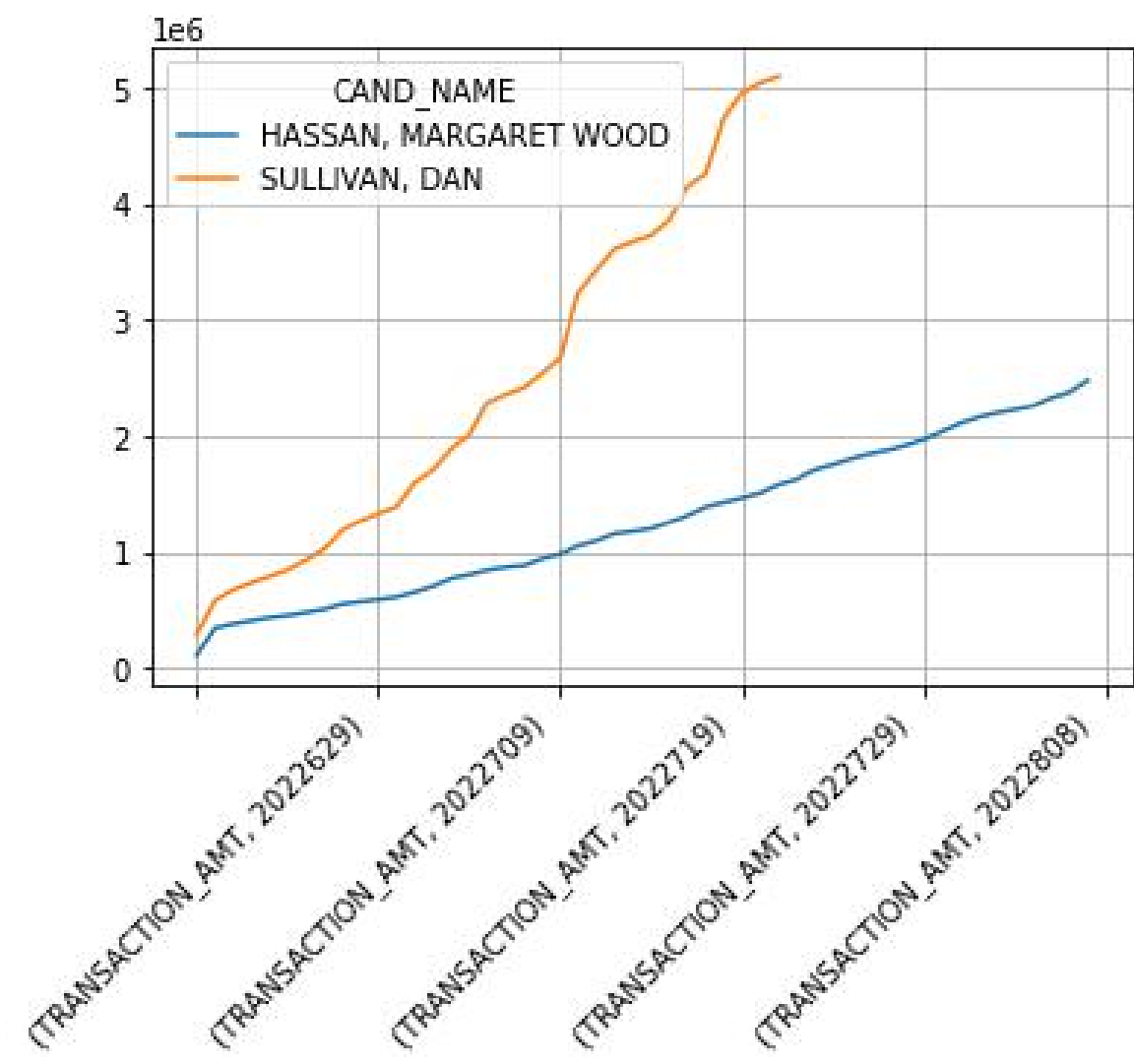
Data Visualization

Sullivan Dan

STATE	TRANSACTION_AMT
TX	480650
FL	447818
CA	422381
MO	349110
NY	297912
CT	291859
WY	265071
VA	230865
GA	208483
IL	177430



Data Visualization



Future Tasks

- 使用k-means聚类对各个州捐款额进行聚类分析;
- 美国贫富差距建模分析;
- 通过选民信息计算可能的选举概率;
-

Share what you learnt~