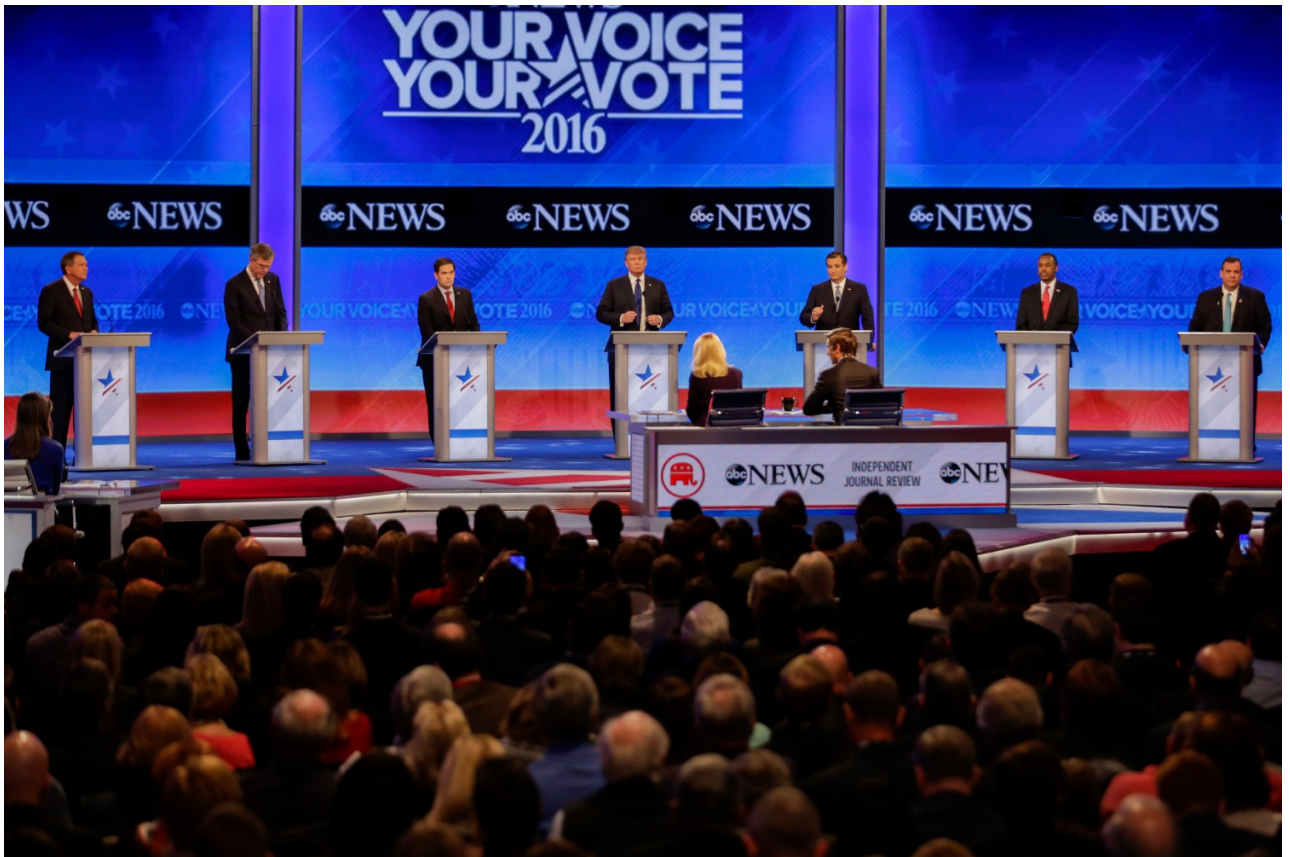


Exploratory Data Analysis of the 2016 Presidential Campaign Expenditures



Importing Data and Libraries

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [181]: campaignExp = pd.read_csv('data/expenditures.csv')
```

Analyzing Campaign Expenditures

```
In [165]: campaignExp['candidate_name'].unique()
```

```
Out[165]: array(['RUBIO, MARCO', 'SANTORUM, RICHARD J.', 'PERRY, JAMES R. (RICK)',
'CARSON, BENJAMIN S.', 'CRUZ, RAFAEL EDWARD 'TED'', 'PAUL, RAND',
'CLINTON, HILLARY RODHAM'], dtype=object)
```

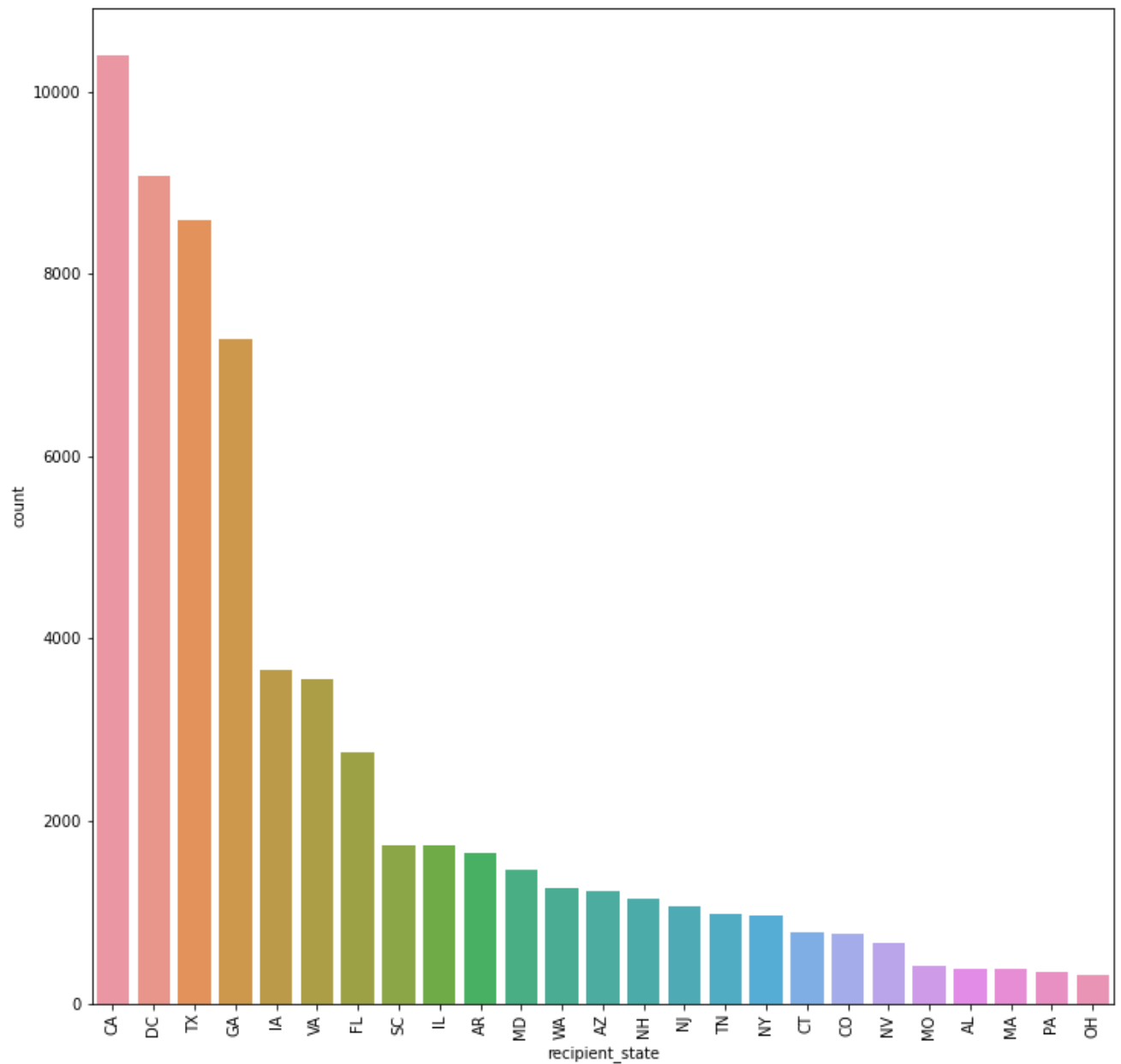
```
In [10]: campaignExp.head()
```

```
Out[10]:
```

	committee_id	candidate_id	candidate_name	recipient_name	disbursement_amount	disbu
0	C00458844	P60006723	RUBIO, MARCO	GREEN, PERRY MR.	500.00	
1	C00458844	P60006723	RUBIO, MARCO	ALASKA REPUBLICAN PARTY	2500.00	
2	C00458844	P60006723	RUBIO, MARCO	JOHNSON, MARK	1000.00	
3	C00458844	P60006723	RUBIO, MARCO	ENTERPRISE	336.09	
4	C00458844	P60006723	RUBIO, MARCO	U.S. SPACE AND ROCKET CENTER	2000.00	

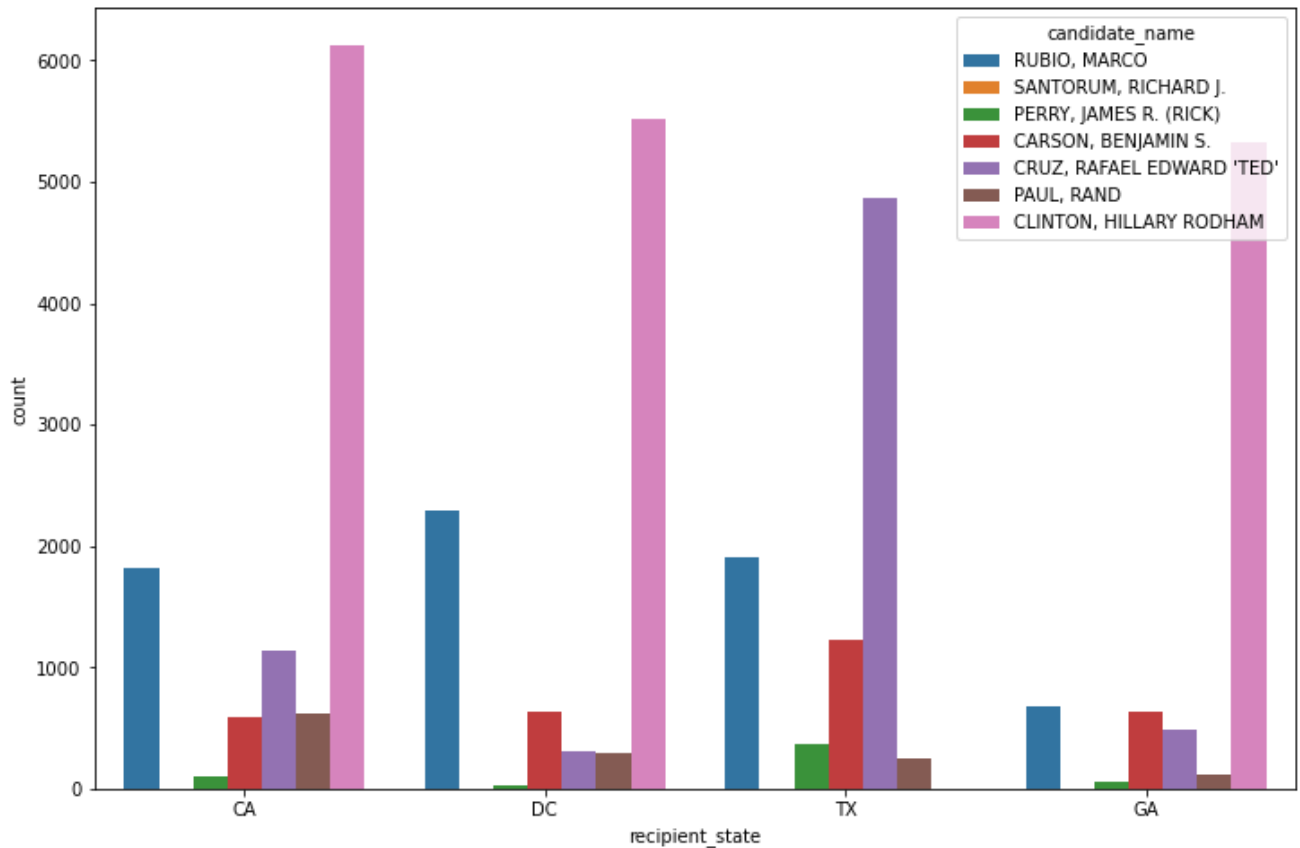
(Q1): What State Did Candidates Spend The Most Money In?

```
In [28]: plt.figure(figsize=(12,12))
sns.countplot(data=campaignExp, x='recipient_state', order = campaignExp['rec
plt.xticks(rotation=90)
plt.show()
```



California, Washington DC, Texas, and Georgia are the most frequent campaign States. Each with over 6000 disbursements

```
In [98]: plt.figure(figsize=(12,8))
sns.countplot(data=campaignExp, x='recipient_state', hue='candidate_name', o
plt.show()
```

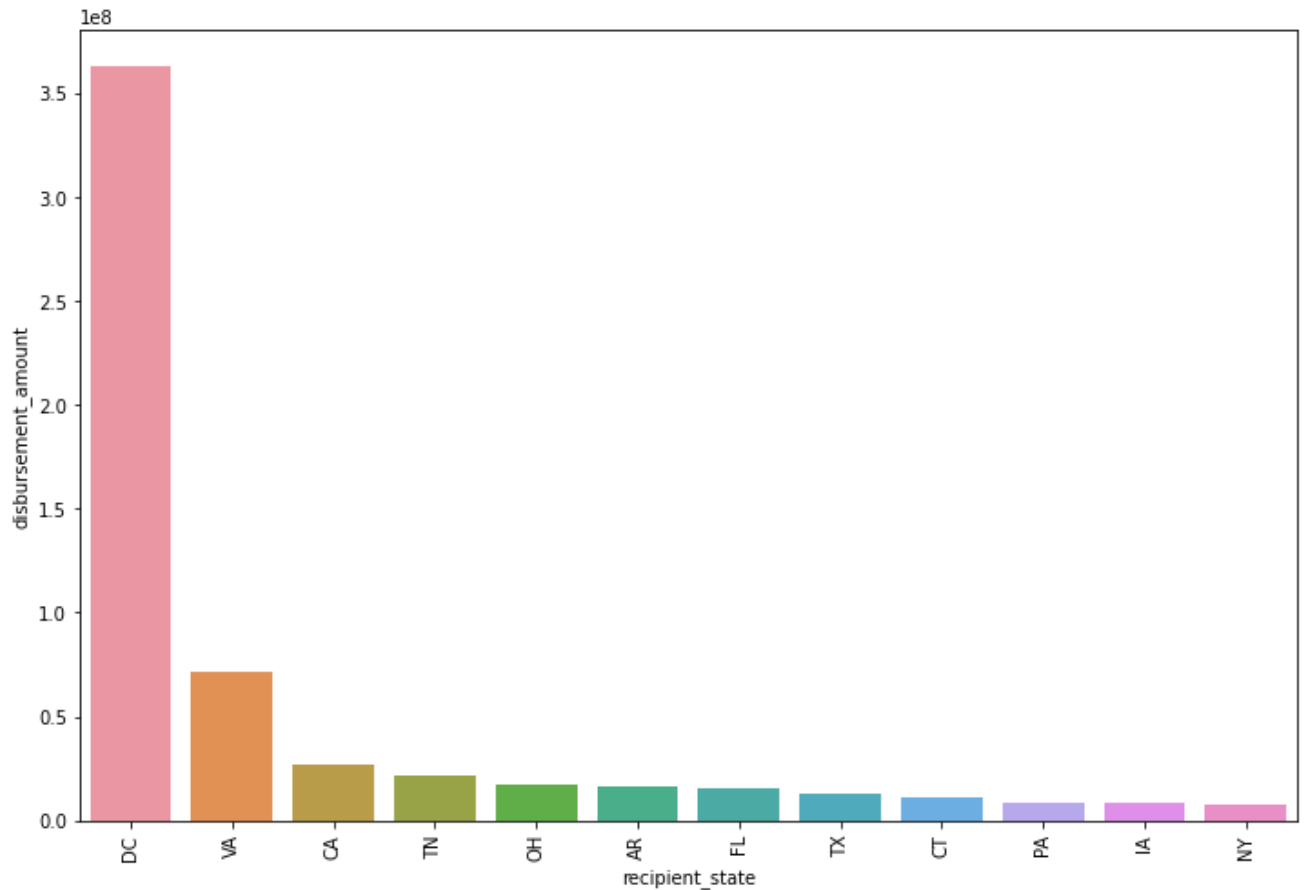


Each State except Texas is where Hilary most frequently spends money for her campaign. Ted cruz spends most significantly in Texas. This is probably because Texas is where Ted Cruz is senator. Marco Rubio consistently is second place in each of the four top states.

(Q2): In Which State is the Most Money Total Spent and by what Candidate?

In [160...

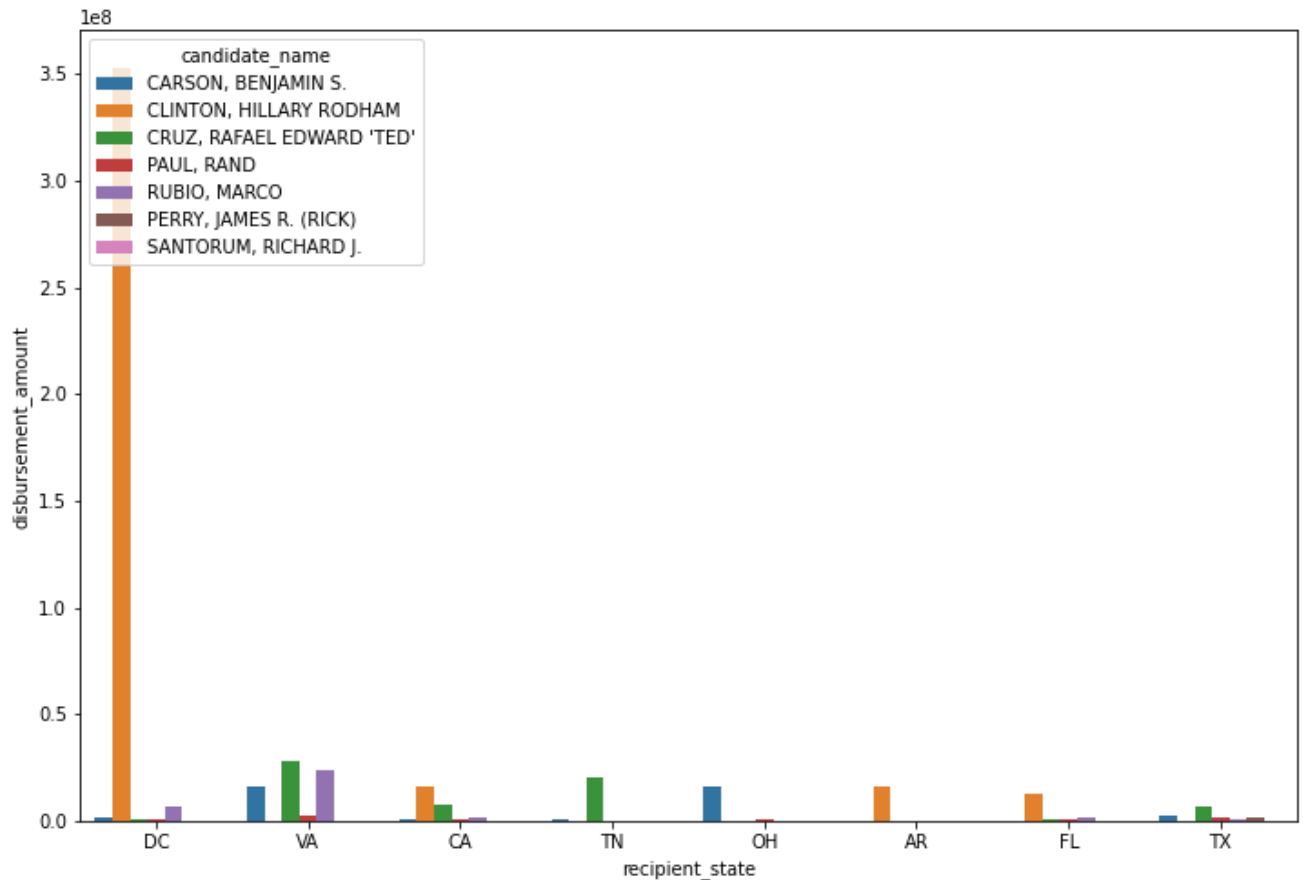
```
campaignExpTotal = campaignExp.groupby(['recipient_state'])['disbursement_amo
plt.figure(figsize=(12,8))
plt.xticks(rotation=90)
sns.barplot(data=campaignExpTotal, x='recipient_state', y='disbursement_amo
plt.show()
```



Washington DC is where the most money is spent despite being second in total disbursements. Virginia is second in total money spent despite being sixth in total recipients. California is in third although they were first in total disbursements.

In [105...

```
campaignExpTotal_cand = campaignExp.groupby(['recipient_state', 'candidate_name'])
plt.figure(figsize=(12,8))
sns.barplot(data=campaignExpTotal_cand, x='recipient_state', y='disbursement_amount')
plt.show()
```



Hilary Clinton spends majority of the donation money in DC, California, Arkansas and Florida, with the rest of the field picking up spending throughout the rest of the states.

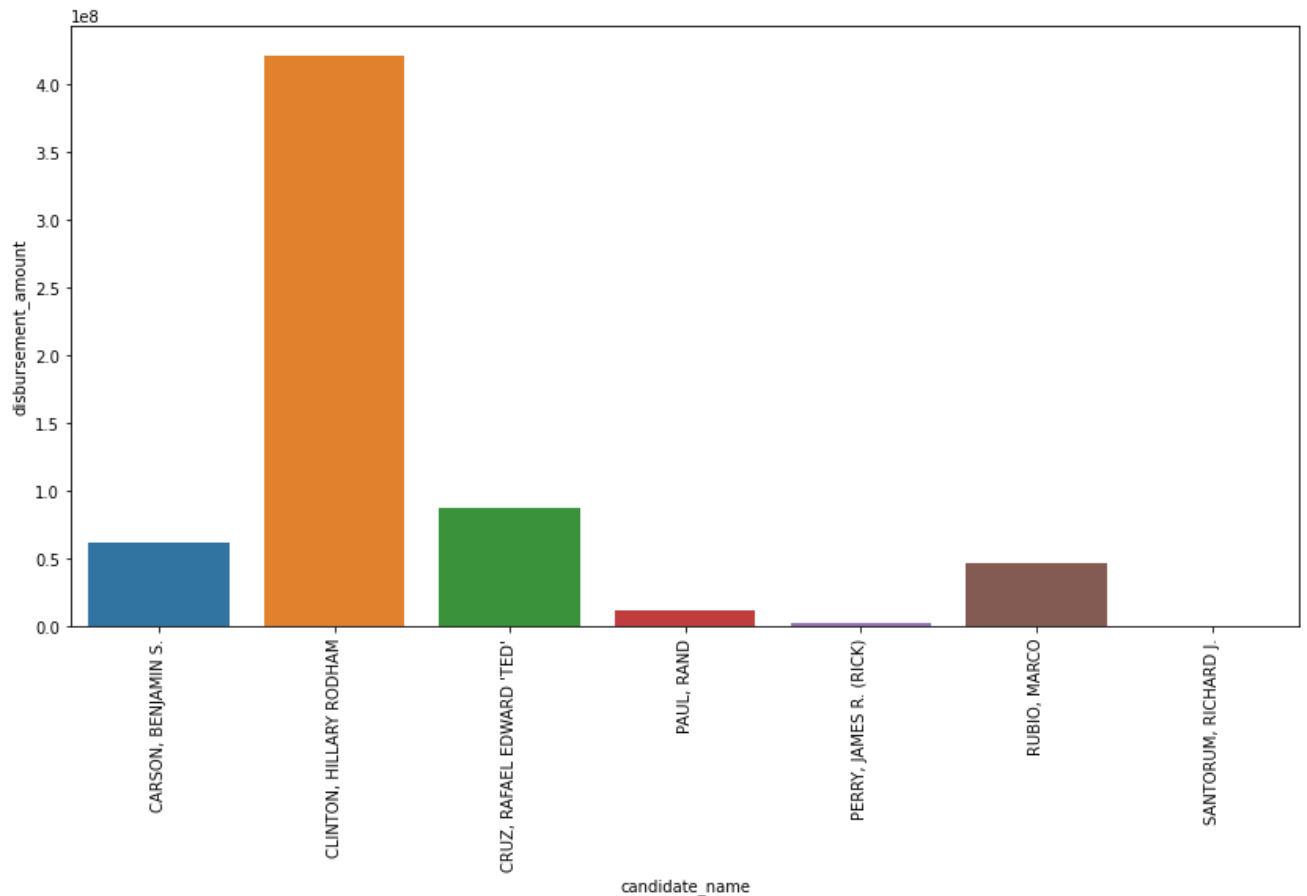
(Q3): Which Candidates spends the most money?

In [155...

```
campaignExpCand = campaignExp.groupby(['candidate_name'])['disbursement_amount']

## Changing Scientific Notation to Normal Values
campaignExpCand['disbursement_amount'] = campaignExpCand['disbursement_amount']
campaignExpCand['disbursement_amount'] = campaignExpCand['disbursement_amount']

plt.figure(figsize=(14,7))
plt.xticks(rotation=90)
sns.barplot(data=campaignExpCand, x='candidate_name', y='disbursement_amount')
plt.show()
```



Hilary Clinton spent the most money, with Ted Cruz and Ben Carson following in respective orders. Richard Santorum was last without any money spent.

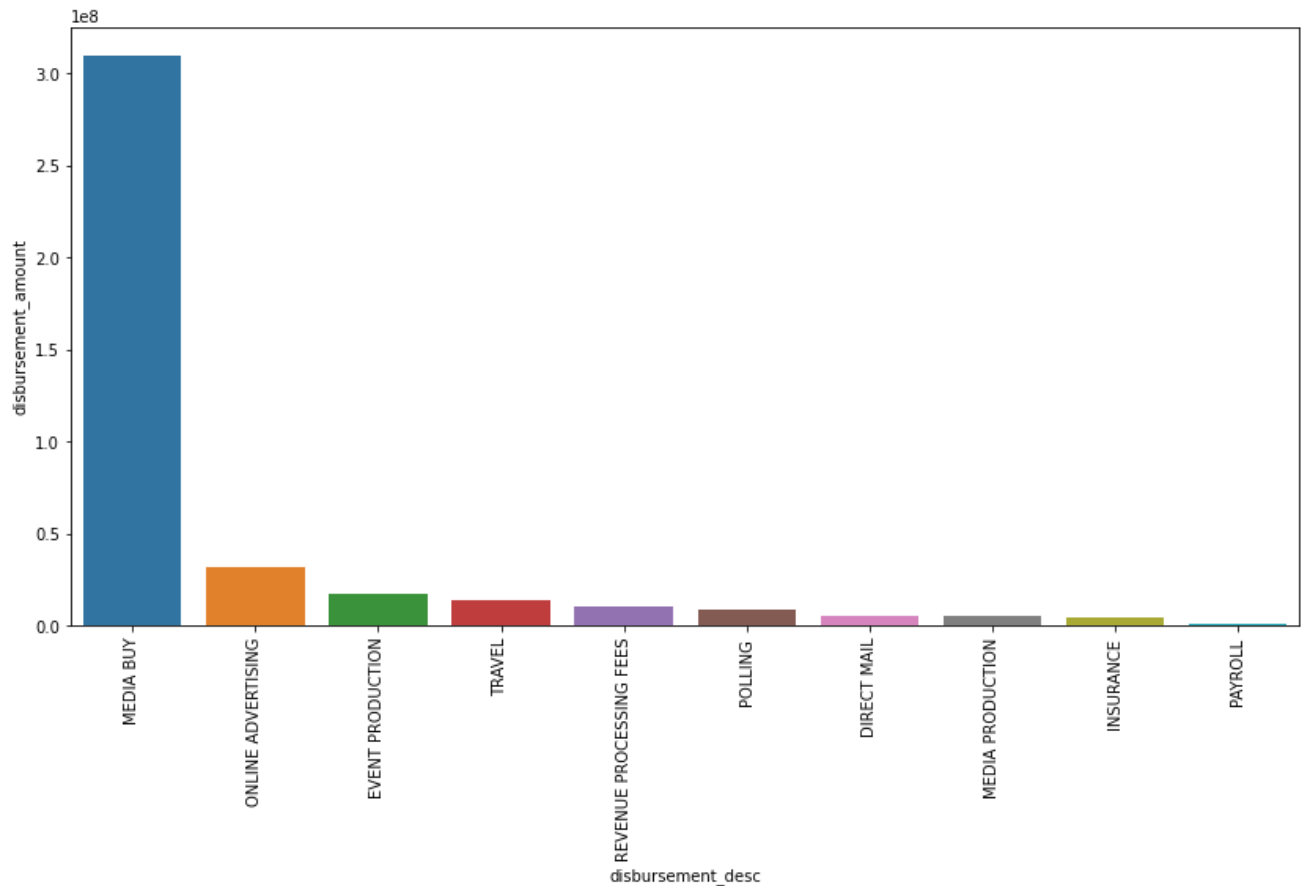
(Q4): What does Hilary Clinton spend most of her campaign donations on?

In [176...

```
campaignExpHilary = campaignExp[campaignExp['candidate_name'] == 'CLINTON, HI
campaignExpHilaryDisb = campaignExpHilary.groupby(['disbursement_desc'])['dis'

plt.figure(figsize=(14,7))
plt.xticks(rotation=90)
sns.barplot(data=campaignExpHilaryDisb, x='disbursement_desc', y='disbursemen

plt.show()
```



Hilary Clinton spent most of her money on Advertising. Mostly TV, with second being online Advertisements.

(Q5): Is there a trend to Hilary Clinton's campaign spending overtime?

In [260...

```
### Changing Disbursement Dates to Date Time
campaignExpHilary['disbursement_date'] = pd.to_datetime(campaignExpHilary['disbursement_date'])

### Adding Month and Year Columns
campaignExpHilary['month_year'] = pd.to_datetime(campaignExpHilary['disbursement_date']).dt.strftime('%m-%Y')

campaignExpHilary.head()
```



```
/var/folders/x1/xynwj5mx66d5r4vf2cjj44q40000gn/T/ipykernel_23565/107008115.py:
2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
campaignExpHilary['disbursement_date'] = pd.to_datetime(campaignExpHilary['disbursement_date'])
```

```
/var/folders/x1/xynwj5mx66d5r4vf2cjj44q40000gn/T/ipykernel_23565/107008115.py:
5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
campaignExpHilary['month_year'] = pd.to_datetime(campaignExpHilary['disbursement_date']).dt.to_period('M')
```

Out [260...

	committee_id	candidate_id	candidate_name	recipient_name	disbursement_amount	c
42318	C00575795	P00003392	CLINTON, HILLARY RODHAM	MARSCH ENTERPRISES	1500.00	
42319	C00575795	P00003392	CLINTON, HILLARY RODHAM	LAUESEN, JENNY	591.99	
42320	C00575795	P00003392	CLINTON, HILLARY RODHAM	TICE, CHARLES	30.00	
42321	C00575795	P00003392	CLINTON, HILLARY RODHAM	GROH, KEVIN DANIEL	1345.63	
42322	C00575795	P00003392	CLINTON, HILLARY RODHAM	MARSCH ENTERPRISES	1500.00	

In [255...

```
campaignExpHilaryTotal_byDate = campaignExpHilary.groupby(['month_year'])['disbursement_amount'].sum()
campaignExpHilary['month_year'] = campaignExpHilary['month_year'].astype(str)

campaignExpHilaryTotal_byDate.head()
```

```
/var/folders/x1/xynwj5mx66d5r4vf2cjj44q40000gn/T/ipykernel_23565/3666238030.py
:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

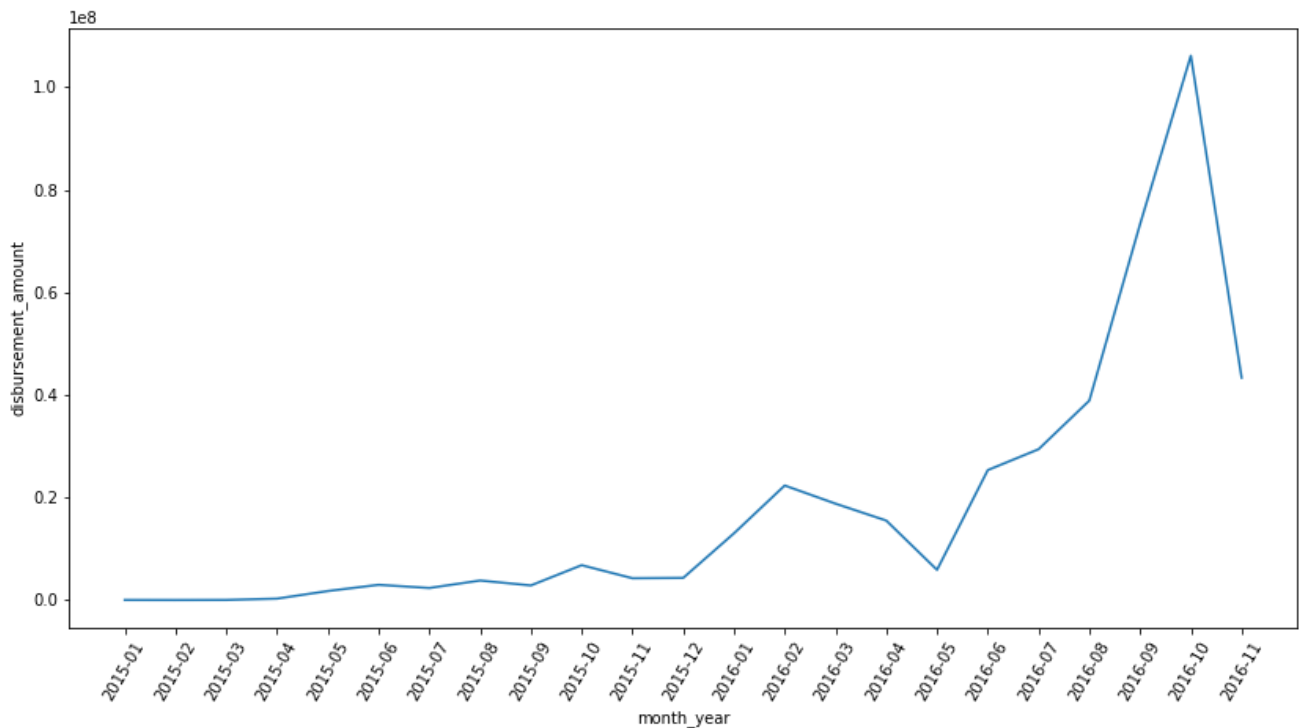
```
campaignExpHilary['month_year'] = campaignExpHilary['month_year'].astype(str)
)
```

Out [255...

	month_year	disbursement_amount
0	2015-01	12499.50
1	2015-02	505.82
2	2015-03	18413.35
3	2015-04	278130.02
4	2015-05	1757662.06

In [258...

```
plt.figure(figsize=(14,7))
plt.xticks(rotation=60)
sns.lineplot(data=campaignExpHilaryTotal_byDate, x='month_year', y='disbursement_amount')
plt.show()
```



In []: