Team 8 Documentation Draft

The Federal Election Commission (FEC) dataset contains information about contributions made to political campaigns and expenditures reported by candidates, parties, political action committees (PACs), and other organizations involved in federal elections in the United States. Let's go through each column and explain its values in short: for more info go here https://www.fec.gov/

1. Definitions

1. col_0: This seems to be an arbitrary column name. It doesn't have a specific description, but it might indicate an index or an identifier for each row in the dataset.

2. Filter_ID: This column might contain an identifier used for filtering or categorizing the data. It could be a unique identifier assigned to different groups or subsets within the dataset.

3. Amendment indicator: Indicates whether the filing is an amendment to a previously filed report. It could have values like "A" for amendment or "N" for no amendment.

4. Report Type: Specifies the type of report being filed, such as quarterly report, pre-election report, post-election report, etc.

5. Election Type: Indicates the type of election to which the contribution or expenditure relates, such as primary, general, special, etc.

6. Office Type: Specifies the type of office for which the candidate is running (e.g., President, Senate, House of Representatives).

7. Transaction Type: Describes the type of transaction being reported, such as contribution, expenditure, loan, etc.

8. Entity Type: Indicates the type of entity making the contribution or expenditure, such as individual, committee, corporation, etc.

9. Contributor: Name of the contributor or entity making the transaction.

10. City: City of the contributor or entity.

11. State: State of the contributor or entity.

12. Zip: ZIP code of the contributor or entity.

13. Employer: Employer of the contributor, if applicable.

14. Occupation: Occupation of the contributor, if applicable.

15. Transaction Date: Date of the transaction.

16. Transaction Amount: Amount of money involved in the transaction.

17. Other ID Number: Another identifier associated with the transaction, possibly provided by the contributor or entity.

18. Transaction ID: Unique identifier for the transaction.

19. Report ID: Identifier for the report in which the transaction is being reported.

20. Memo Code: A code indicating the purpose or nature of the transaction, usually used for additional information.

21. Memo Text: Additional text providing details or explanations regarding the transaction.

22. FEC Record Number: Unique identifier assigned by the FEC to each transaction record.

Each column provides specific details about the contributions or expenditures reported, helping to track and analyze financial activities related to federal elections.

The House data Definition

The "1976_2022 house dataset" appears to be a dataset containing information about U.S. House of Representatives elections from 1976 to 2022.

1. candidatevotes:
   - Data Type: Integer
   - Description: The number of votes received by a specific candidate in a particular election.

2. totalvotes:

- Data Type: Integer
- Description: The total number of votes cast in the election for the specific race or district.

3. version:
- Data Type: Date (unparsed)
- Description: A version identifier or timestamp indicating when the dataset was last updated or modified.

4. fusion_ticket:
- Data Type: Boolean
- Description: Indicates whether the candidate was running on a fusion ticket, which typically refers to a candidate being endorsed by multiple political parties.

5. state_ic:
- Data Type: Integer
- Description: A numeric identifier for the state or jurisdiction in which the election took place.

6. office:
- Data Type: Text
- Description: The office or position being sought in the election, which in this case would be the U.S. House of Representatives.

7. district:
- Data Type: Integer
- Description: The district number or identifier for the specific House seat being contested in the election.

8. stage:
- Data Type: Text
- Description: Indicates the stage of the election process, which could include primary, general, or other stages.

9. runoff:
- Data Type: Boolean
- Description: Indicates whether the election resulted in a runoff between the top candidates, typically occurring when no candidate receives a majority of the votes in the initial election.

10. special:
   - Data Type: Boolean
   - Description: Indicates whether the election is a special election held outside of the regular election cycle to fill a vacant House seat.

11. candidate:
   - Data Type: Text
   - Description: The name of the candidate running for the House seat.

12. party:
   - Data Type: Text
   - Description: The political party affiliation of the candidate.

13. writein:
   - Data Type: Boolean
   - Description: Indicates whether the candidate's name was written in by voters rather than appearing on the official ballot.

14. state_po:
   - Data Type: Text
   - Description: This field represents the two-letter postal abbreviation for the state in which the election took place. For example, "CA" for California or "NY" for New York. This abbreviation follows the standard postal abbreviations used in the United States.

15. mod:
   - Data Type: Unknown (Assumed to be Text or Integer)
   - Description: The "mod" field doesn't have a clear definition based on the provided information. It may refer to a modification or a specific code related to the election data. Without further context or documentation, it's challenging to provide a precise explanation of this field.

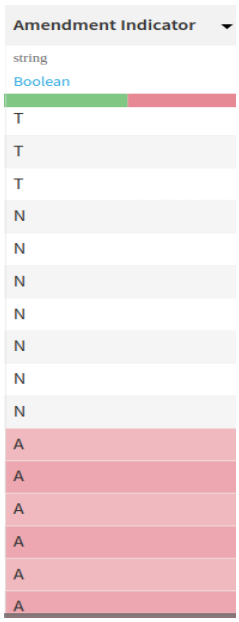
16. unofficial:
   - Data Type: Boolean
   - Description: This field indicates whether the election results are unofficial. Unofficial results are typically reported immediately after the election and may not yet be certified by the relevant election authorities. Once results are certified, they become official. Boolean values in this field can be interpreted as "True" if the results are unofficial and "False" if they are official.

2. Cleaning

1. 2008 Data

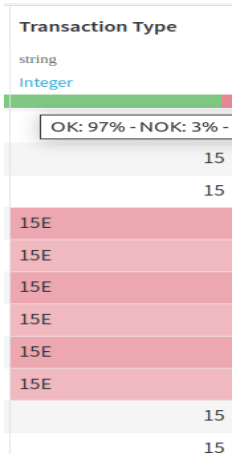
- **Amendment Indicator**

For boolean values, it should be True or false. But currently, it is a mix of T, N, and A. So converted this to a string type and removed the T type since they are not required according to the docs.



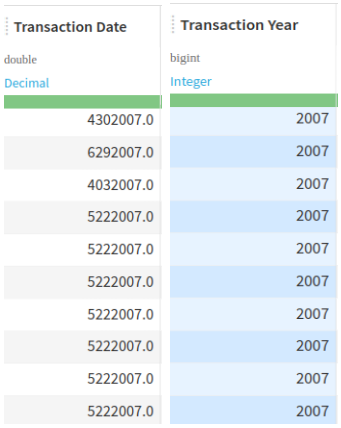
- **Transaction Type**

This was a mix of numerical values and Alpha-numeric values thus we converted them to a constant String(Storage Data) and Text (Meaning of the Data).



- **Transaction Date**

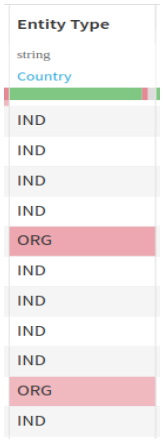
This was represented as a Decimal numerical values thus we converted them to bigint/Integer value, then we extracted only the Year, because the Month could be determined from the Report Type or could be approximated. The date was not important at all. Therefore, we also renamed the column from Transaction Date to Transaction Year.



Transaction Date double Decimal	Transaction Year bigint Integer
4302007.0	2007
6292007.0	2007
4032007.0	2007
5222007.0	2007
5222007.0	2007
5222007.0	2007
5222007.0	2007
5222007.0	2007
5222007.0	2007
5222007.0	2007

- **Entity Type**

This was being labeled as countries , while it was the contributors type , thus it was returning invalid values. This was changed to String/Text format as part of the cleaning process.

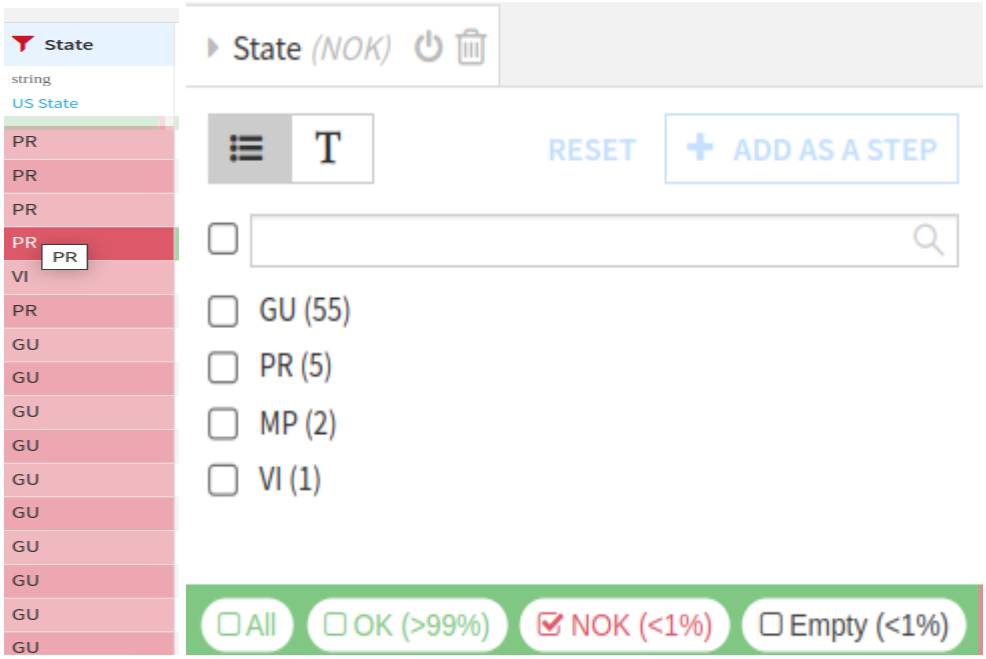


Entity Type string Country
IND
IND
IND
IND
ORG
IND
IND
IND
IND
ORG
IND

- **State**

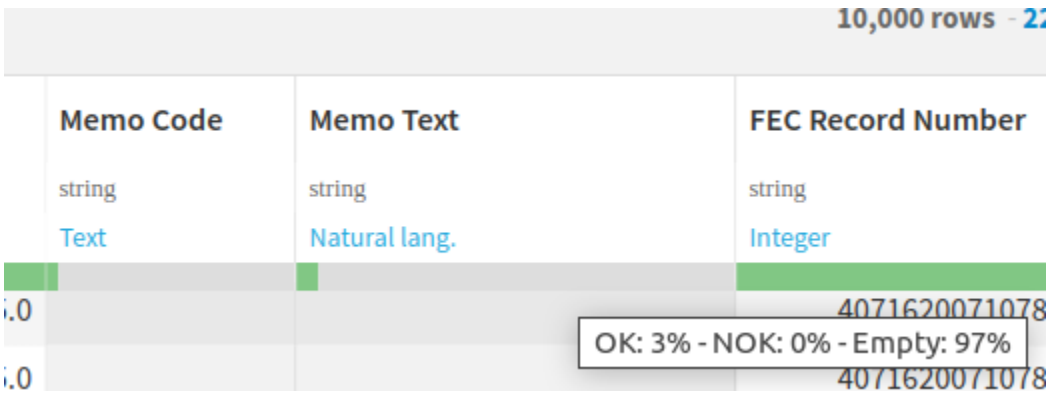
PR, MP,VI and GU are not invalid in the context of the United States, but rather they represent different regions or territories. But due to their small size and low impact in the overall dataset we decided to remove their corresponding rows entirely.

1. For instance, PR stands for Puerto Rico, which is a U.S. territory rather than a state. While residents of Puerto Rico are U.S. citizens, Puerto Rico itself is not a U.S. state but rather an unincorporated territory.



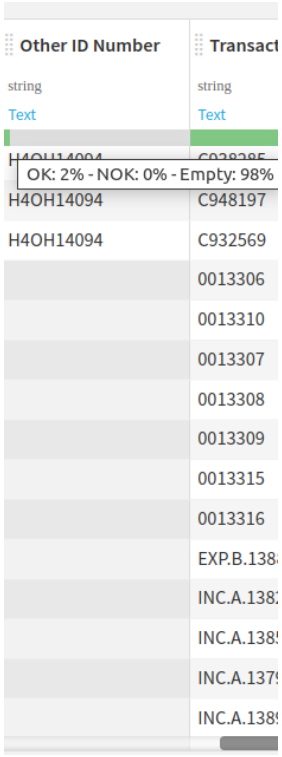
- **Memo Code and Memo Text**

We decided to remove these two columns because 97% of their data were missing , rendering them irrelevant for any further analysis or visualizationvisualization.

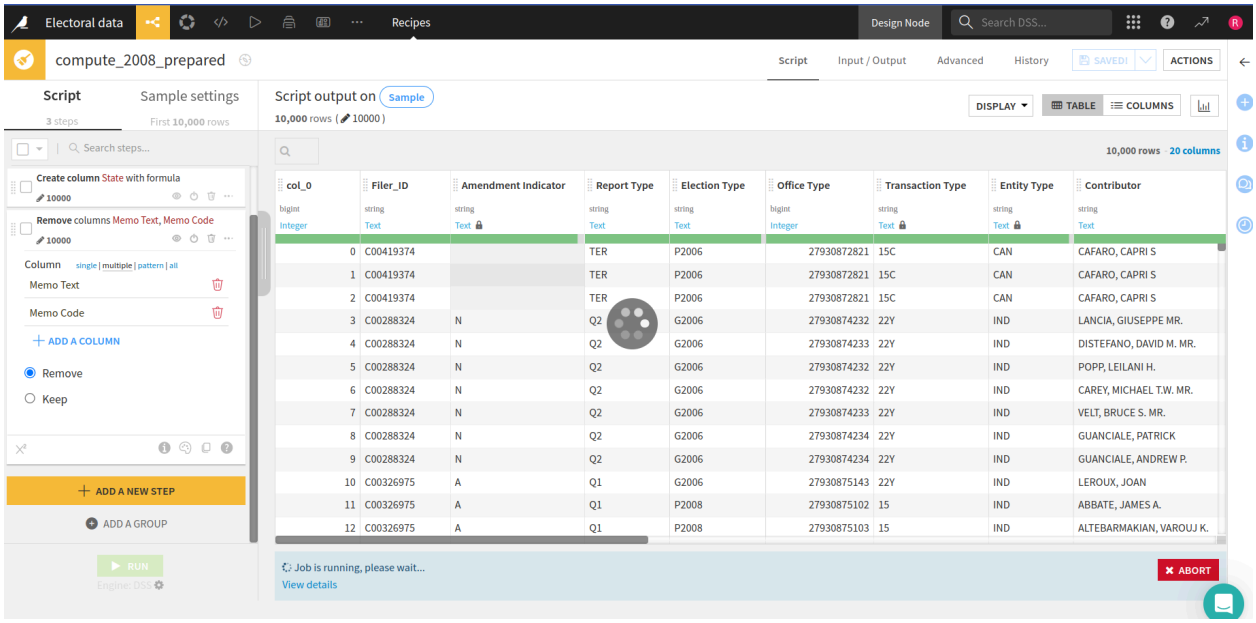


- **Other ID Number**

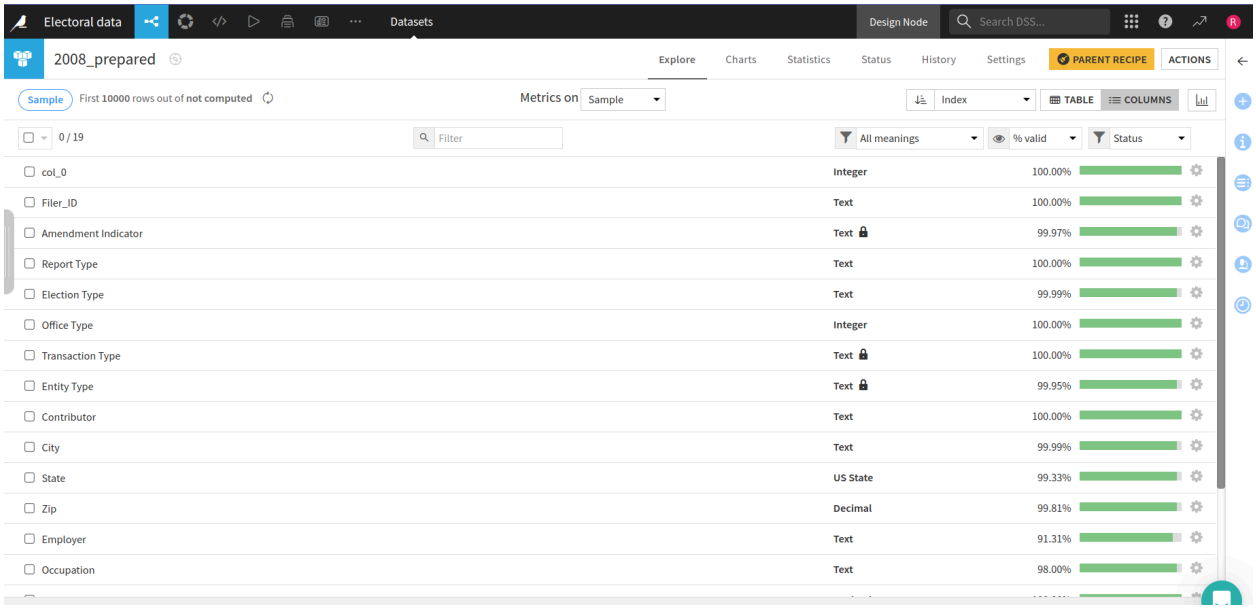
We decided to remove this column because 98% of it's data were missing , rendering the column irrelevant for any further analysis or visualization.

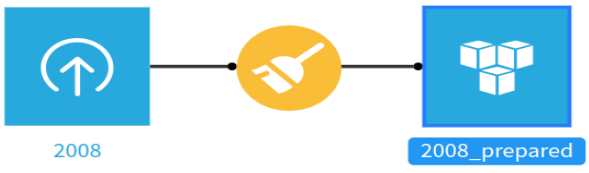


Cleaning Job processing



Cleaning Job completed.





2. 2010 Data

- **Amendment Indicator: Applied the same operation as 2008 Data.**
- **Transaction Type: Applied the same operation as 2008 Data.**
- **Transaction Date: Applied the same operation as 2008 Data.**
- **Entity Type: Applied the same operation as 2008 Data.**
- **Memo Code: Applied the same operation as 2008 Data.**
- **Memo Text: Applied the same operation as 2008 Data.**
- **Other ID Number: Applied the same operation as 2008 Data.**

3. 2012 Data

- **Amendment Indicator: Applied the same operation as 2008 Data.**
- **Transaction Type: Applied the same operation as 2008 Data.**
- **Transaction Date: Applied the same operation as 2008 Data.**
- **Entity Type: Applied the same operation as 2008 Data.**
- **State: Applied the same operation as 2008 Data.**
- **Memo Code: Applied the same operation as 2008 Data.**
- **Memo Text: Applied the same operation as 2008 Data.**
- **Other ID Number: Applied the same operation as 2008 Data.**
- **Election Type: This was being parsed as Date while it should be Text format. This was processed accordingly and converted to Text format.**

4. 2014 Data

- **Amendment Indicator: Applied the same operation as 2008 Data.**
- **Transaction Type: Applied the same operation as 2008 Data.**
- **Transaction Date: Applied the same operation as 2008 Data.**
- **Entity Type: Applied the same operation as 2008 Data.**
- **State: Applied the same operation as 2008 Data.**
- **Memo Code: Applied the same operation as 2008 Data.**
- **Memo Text: Applied the same operation as 2008 Data.**
- **Other ID Number: Applied the same operation as 2008 Data.**
- **Transaction ID: This was being parsed as Integer ; so we processed it to become Text format.**

5. 2016 Data

- **Amendment Indicator: Applied the same operation as 2008 Data.**
- **Transaction Type: Applied the same operation as 2008 Data.**
- **Transaction Date: Applied the same operation as 2008 Data.**
- **Entity Type: Applied the same operation as 2008 Data.**
- **State: Applied the same operation as 2008 Data.**
- **Memo Code: Applied the same operation as 2008 Data.**
- **Memo Text: Applied the same operation as 2008 Data.**
- **Other ID Number: Applied the same operation as 2008 Data.**

6. 2018 Data

- **Amendment Indicator: Applied the same operation as 2008 Data.**
- **Transaction Type: Applied the same operation as 2008 Data.**
- **Transaction Date: Applied the same operation as 2008 Data.**
- **Entity Type: Applied the same operation as 2008 Data.**
- **State: Applied the same operation as 2008 Data.**
- **Memo Code: Applied the same operation as 2008 Data.**
- **Memo Text: Applied the same operation as 2008 Data.**
- **Other ID Number: Applied the same operation as 2008 Data.**

7. 2020_a Data

- **Transaction Type: Applied the same operation as 2008 Data.**
- **Transaction Date: Applied the same operation as 2008 Data.**
- **Entity Type: Applied the same operation as 2008 Data.**
- **Memo Code: Applied the same operation as 2008 Data.**
- **Memo Text: Applied the same operation as 2008 Data.**
- **Other ID Number: Applied the same operation as 2008 Data.**

8. 2020_b Data

- **Amendment Indicator: Applied the same operation as 2008 Data.**
- **Transaction Type: Applied the same operation as 2008 Data.**
- **Transaction Date: Applied the same operation as 2008 Data.**
- **Entity Type: Applied the same operation as 2008 Data.**
- **Zip: Remove wrong zip code (had a Text Entry).**
- **Memo Code: Applied the same operation as 2008 Data.**
- **Memo Text: Applied the same operation as 2008 Data.**
- **Other ID Number: Applied the same operation as 2008 Data.**

9. 2020_c Data

- **Amendment Indicator: Applied the same operation as 2008 Data.**
- **Transaction Type: Applied the same operation as 2008 Data.**
- **Transaction Date: Applied the same operation as 2008 Data.**
- **Entity Type: Applied the same operation as 2008 Data.**
- **Memo Code: Applied the same operation as 2008 Data.**
- **Memo Text: Applied the same operation as 2008 Data.**
- **Other ID Number: Applied the same operation as 2008 Data.**

10. 2022_a Data

- **Amendment Indicator: Applied the same operation as 2008 Data.**
- **Transaction Type: Applied the same operation as 2008 Data.**
- **Transaction Date: Applied the same operation as 2008 Data.**
- **Entity Type: Applied the same operation as 2008 Data.**
- **Memo Code: Applied the same operation as 2008 Data.**
- **Memo Text: Applied the same operation as 2008 Data.**
- **Other ID Number: Applied the same operation as 2008 Data.**
- **Transaction ID: This was being parsed as Integer ; so we processed it to become Text format.**

11. 2022_b Data

- **Amendment Indicator: Applied the same operation as 2008 Data.**
- **Transaction Type: Applied the same operation as 2008 Data.**
- **Transaction Date: Applied the same operation as 2008 Data.**
- **Entity Type: Applied the same operation as 2008 Data.**
- **Memo Code: Applied the same operation as 2008 Data.**
- **Memo Text: Applied the same operation as 2008 Data.**
- **Other ID Number: Applied the same operation as 2008 Data.**

3. Visualization

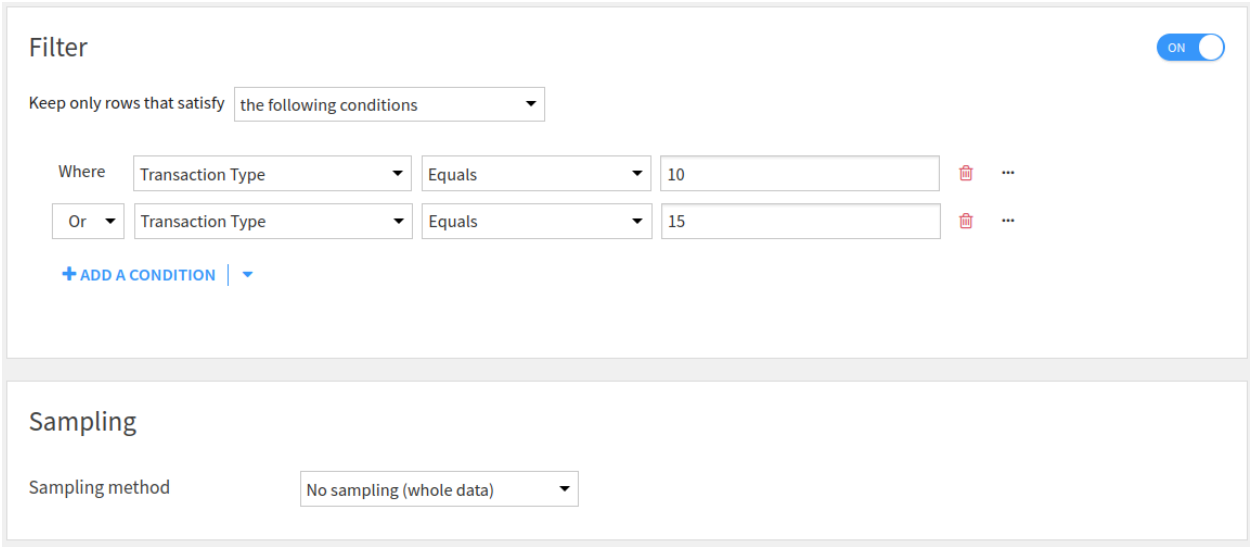
1. We would like to visualize the we only want individual contributions to candidates that are monetary donations instead of in-kind ?

We have selected the Transaction Type 10 and 15.

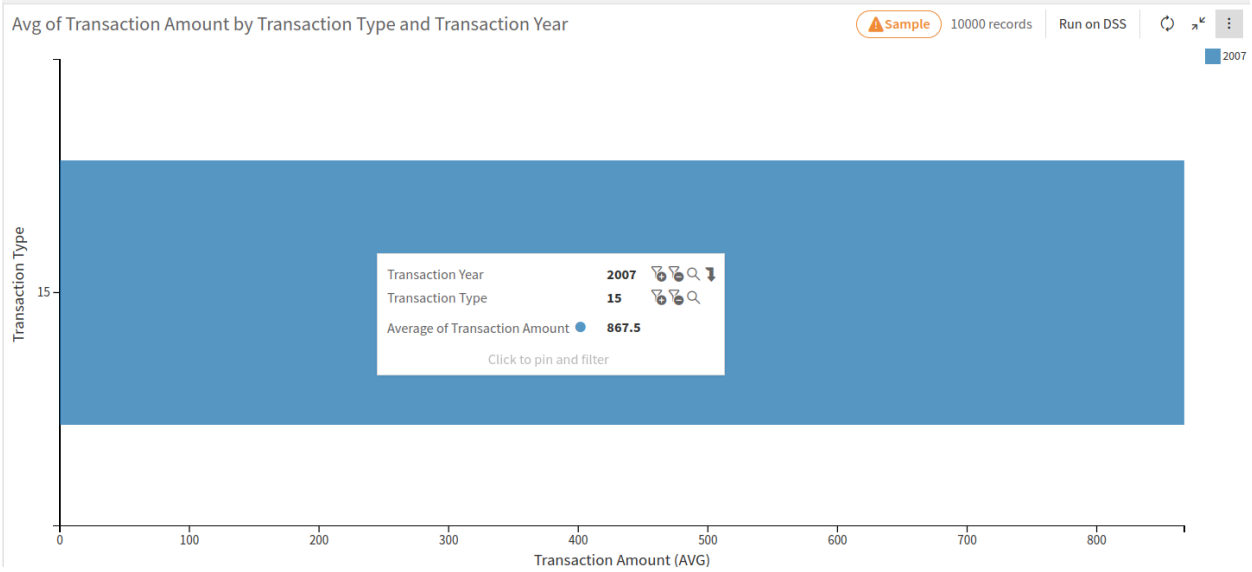
Transaction type	Transaction type description
10	Contribution to Independent Expenditure-Only Committees (Super PACs), Political Committees with non-contribution accounts (Hybrid PACs) and nonfederal party "soft money" accounts (1991-2002) from a person (individual, partnership, limited liability company, corporation, labor organization, or any other organization or group of persons)
15	Contribution to political committees (other than Super PACs and Hybrid PACs) from an individual, partnership or limited liability company

1. 2008 Data

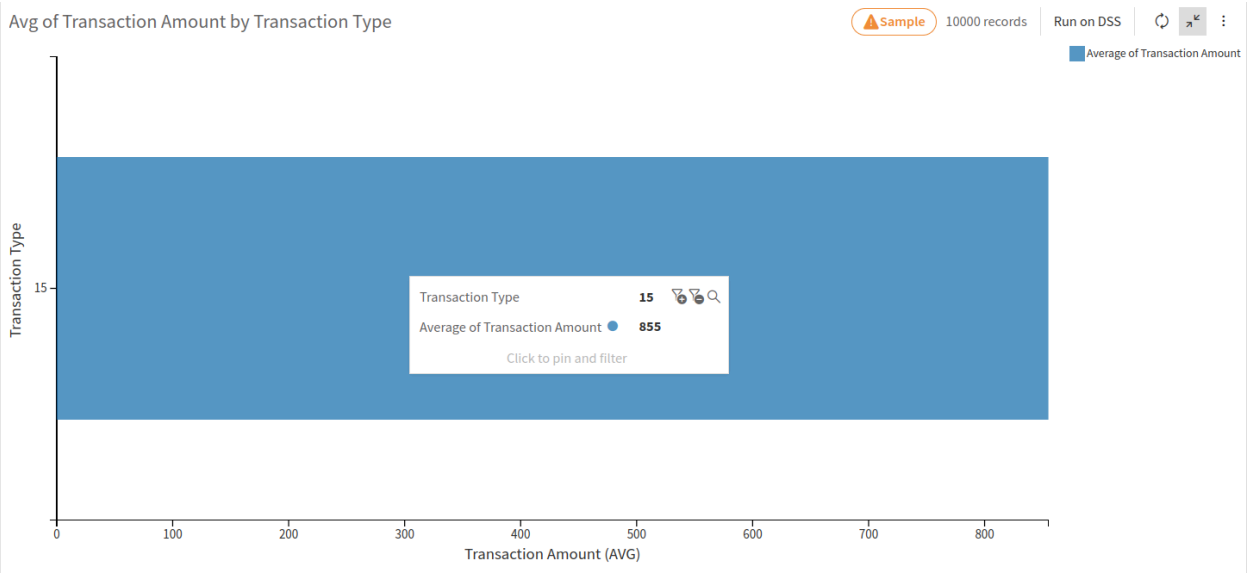
Filter the data (Same Operation has been applied to subsequent Datasets.)



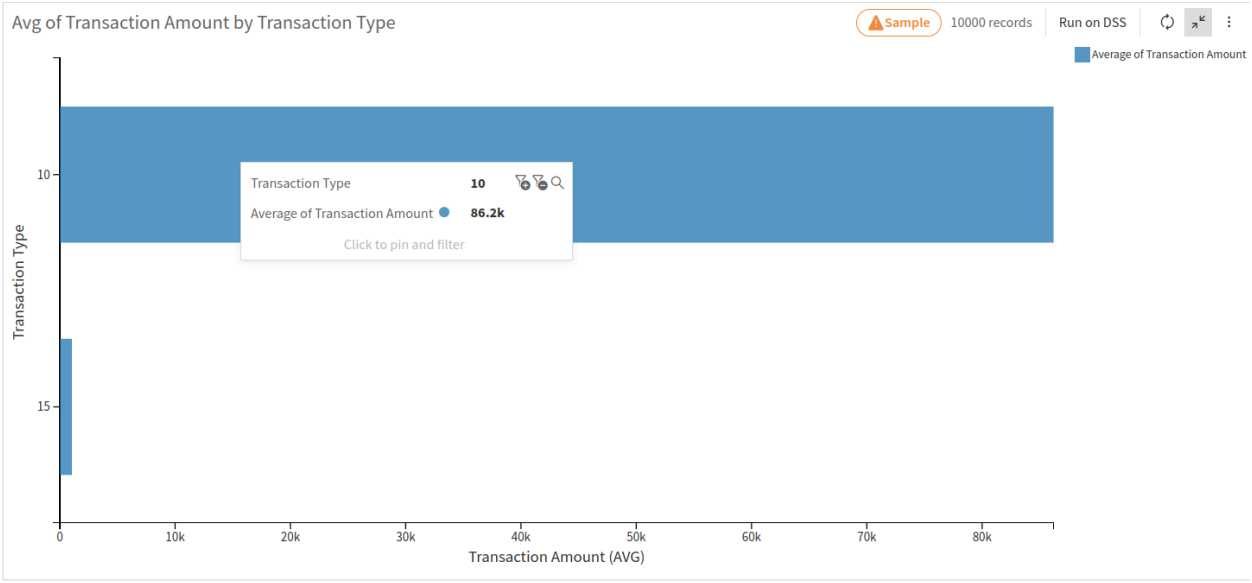
Data Visualization

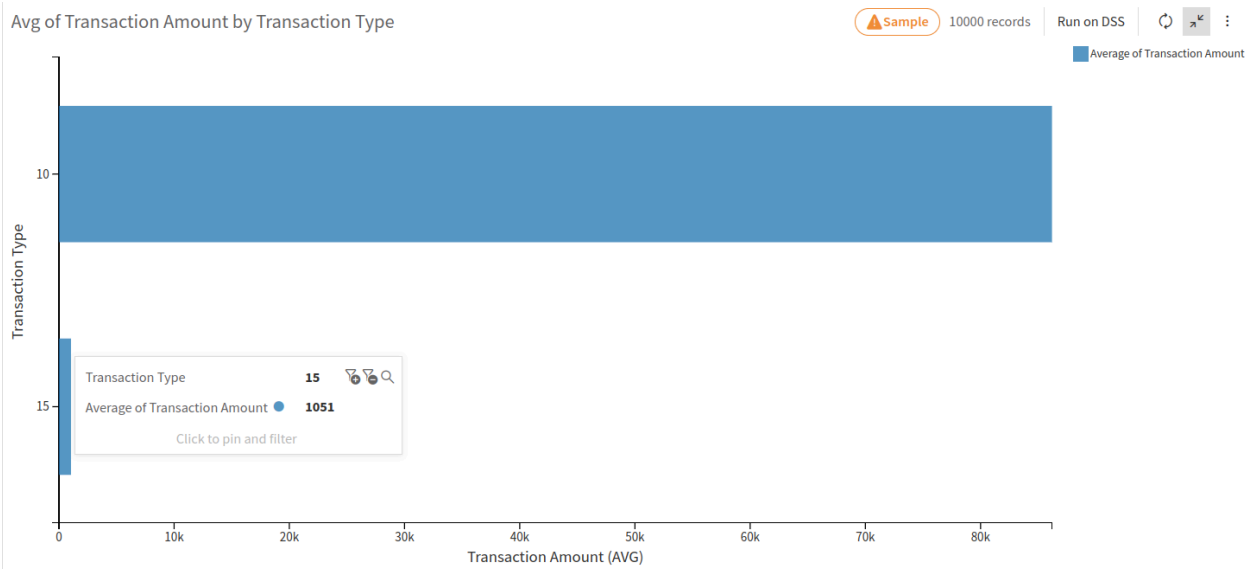


2. 2010 Data

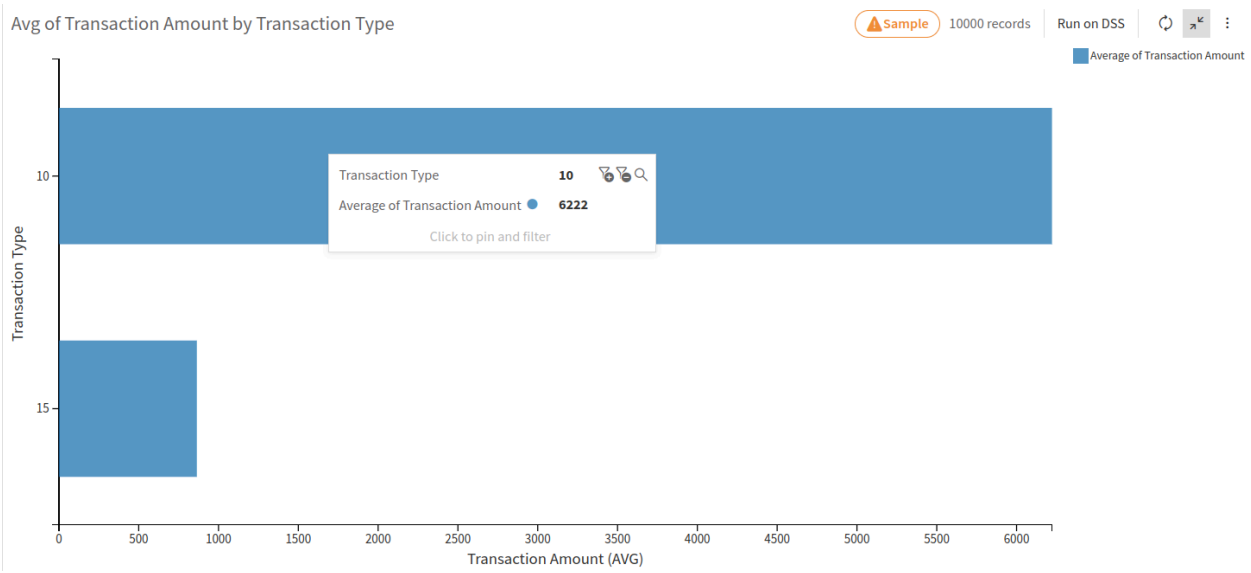

Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
15
Transaction Type
15
Average of Transaction Amount
855
Click to pin and filter
0
100
200
300
400
500
600
700
800
Transaction Amount (AVG)

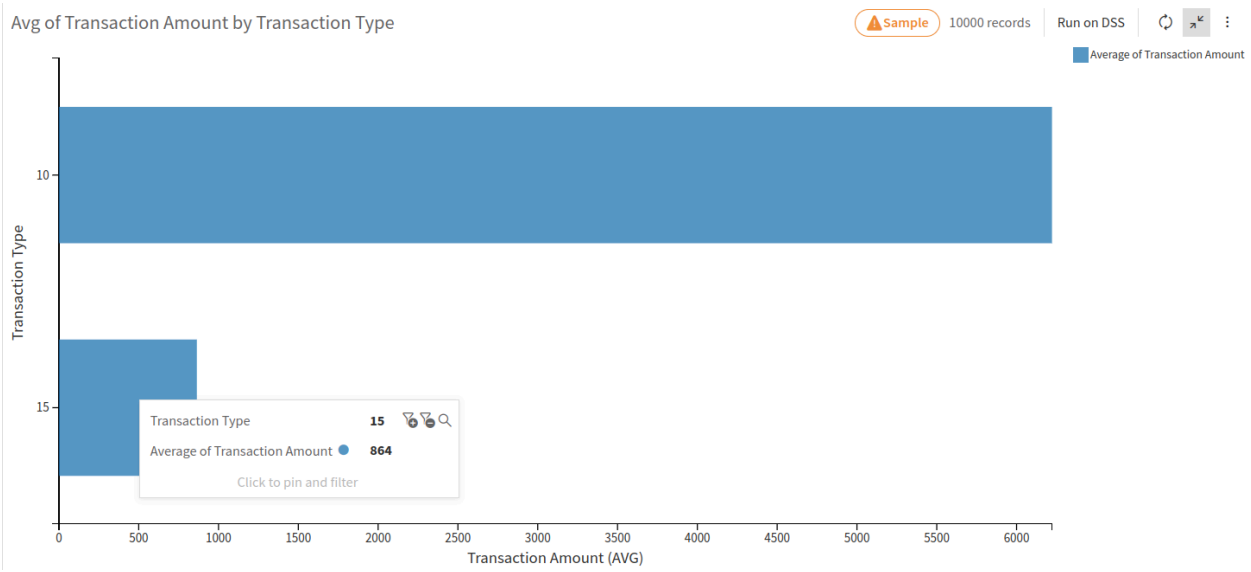
3. 2012 Data


Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
15
Transaction Type
10
Average of Transaction Amount
86.2k
Click to pin and filter
0
10k
20k
30k
40k
50k
60k
70k
80k
Transaction Amount (AVG)

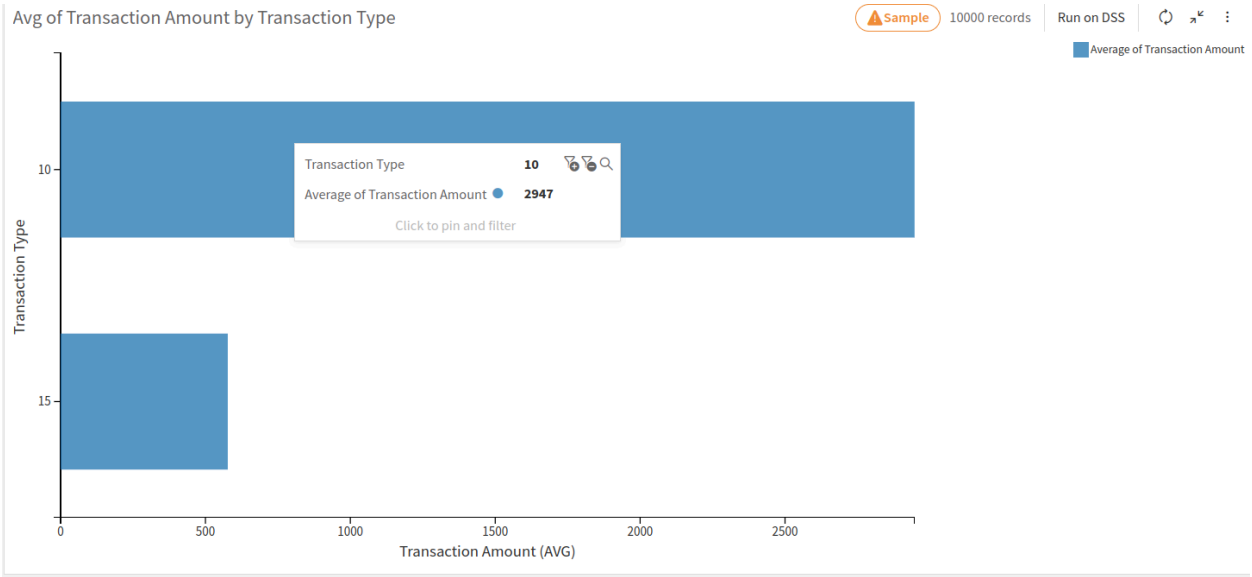

Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
15
Transaction Type
15
Average of Transaction Amount
1051
Click to pin and filter
0
10k
20k
30k
40k
50k
60k
70k
80k
Transaction Amount (AVG)

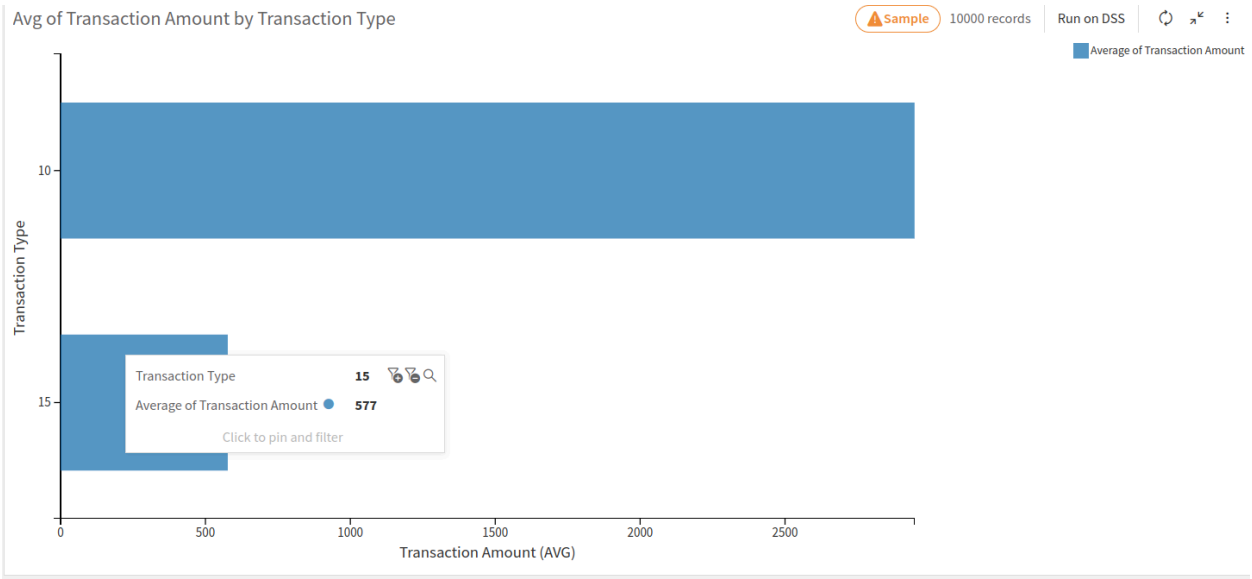
4. 2014 Data


Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
15
Transaction Type 10
Average of Transaction Amount 6222
Click to pin and filter
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
Transaction Amount (AVG)

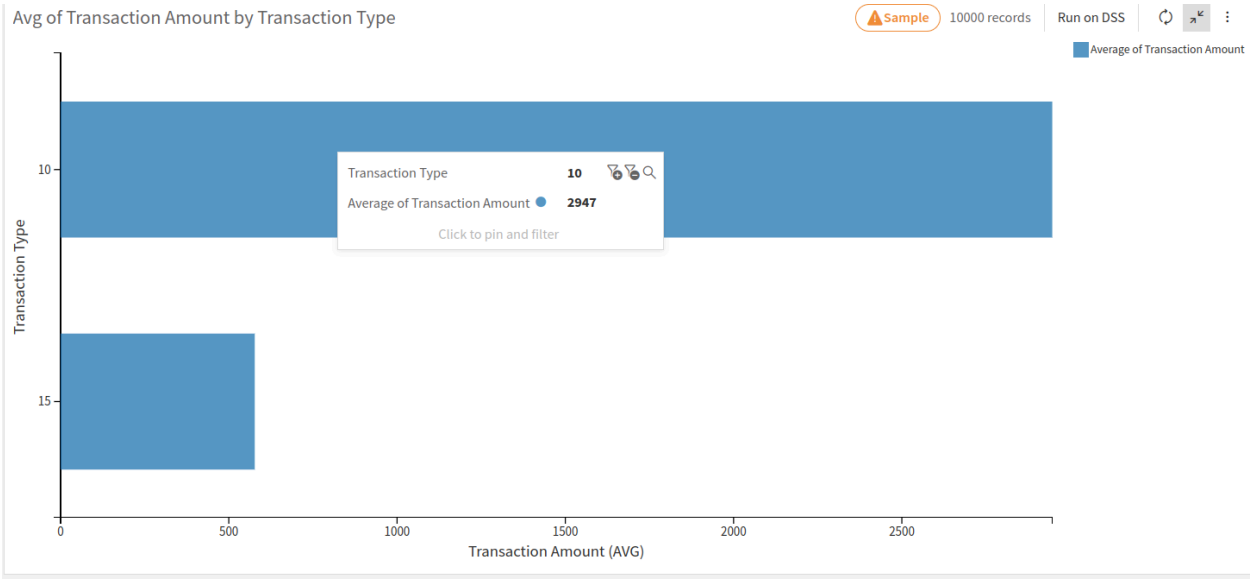

Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
15
Transaction Type 15
Average of Transaction Amount 864
Click to pin and filter
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
Transaction Amount (AVG)

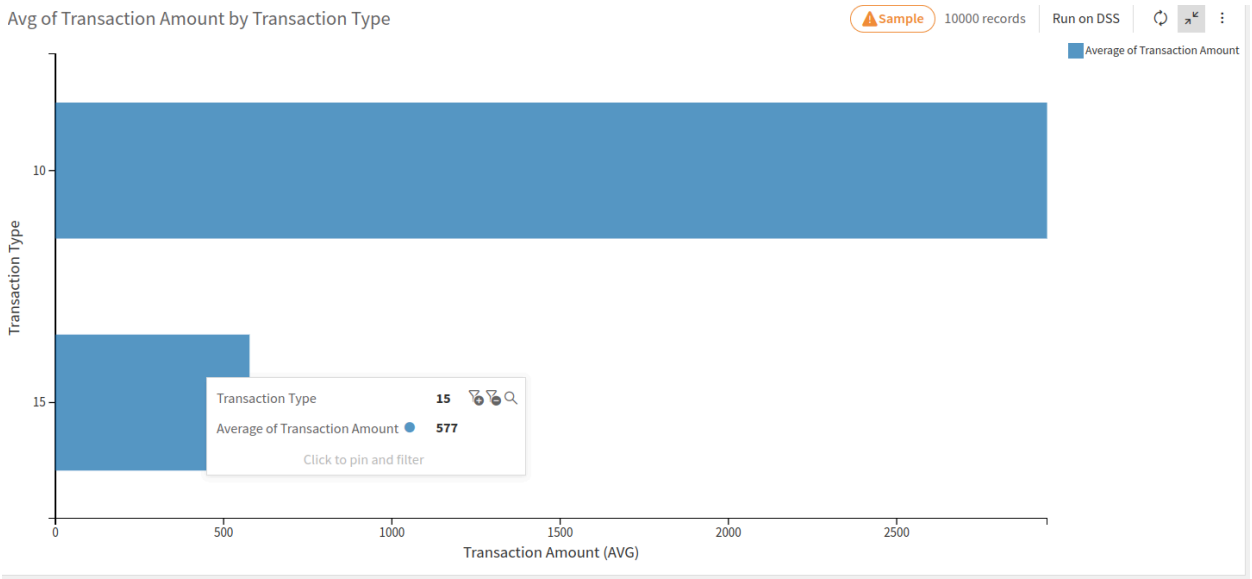
5. 2016 Data


Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
15
0
500
1000
1500
2000
2500
Transaction Amount (AVG)
Transaction Type 10
Average of Transaction Amount 2947
Click to pin and filter

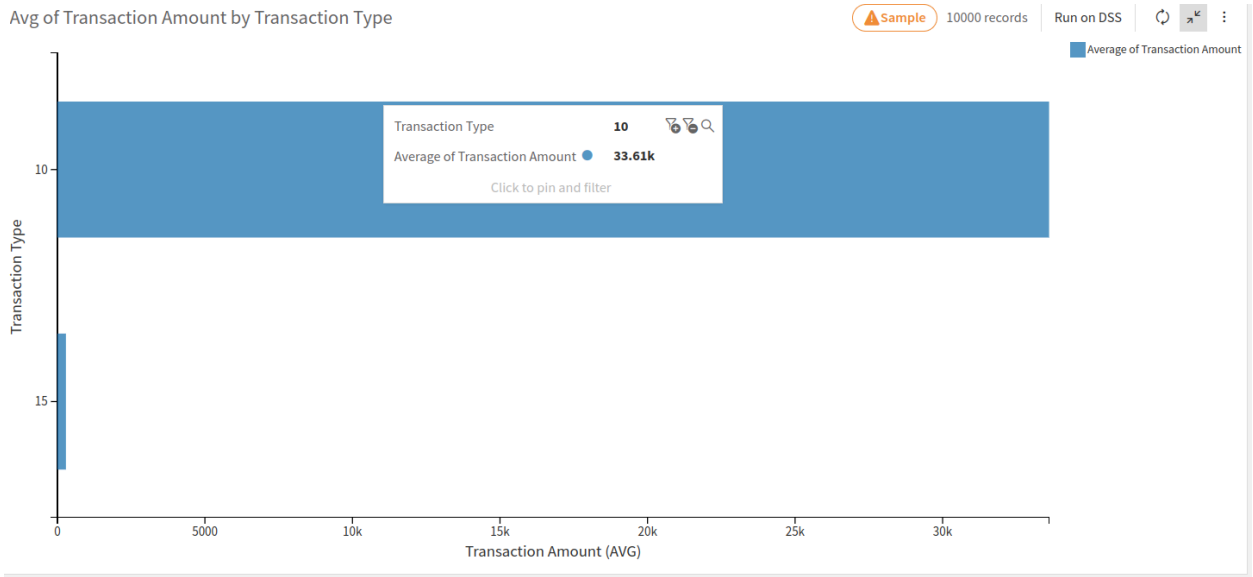

Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
15
0
500
1000
1500
2000
2500
Transaction Amount (AVG)
Transaction Type 15
Average of Transaction Amount 577
Click to pin and filter

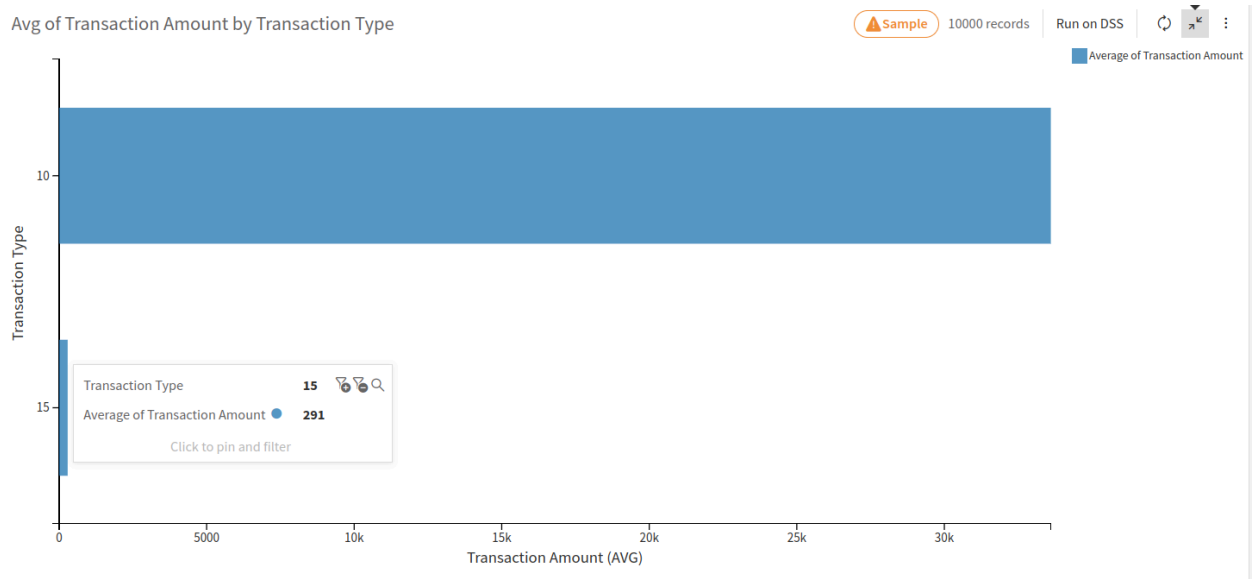
6. 2018 Data


Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
15
Transaction Type 10
Average of Transaction Amount 2947
Click to pin and filter
0
500
1000
1500
2000
2500
Transaction Amount (AVG)

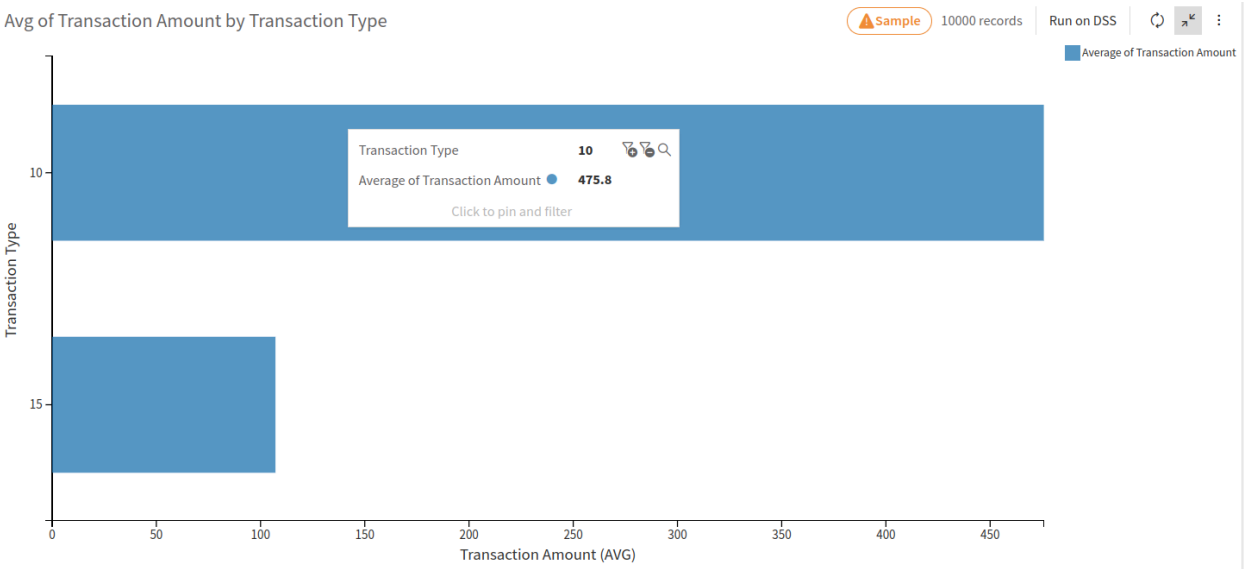

Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
15
Transaction Type 15
Average of Transaction Amount 577
Click to pin and filter
0
500
1000
1500
2000
2500
Transaction Amount (AVG)

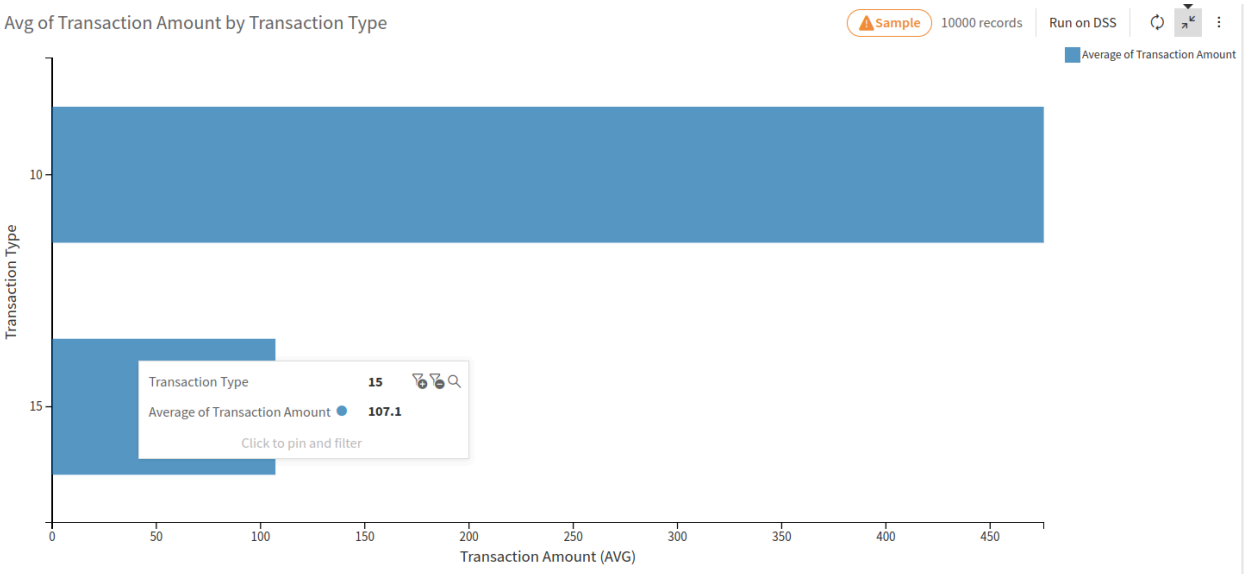
7. 2020_a Data


Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
Average of Transaction Amount
33.61k
Click to pin and filter
Transaction Type
10
15
0
5000
10k
15k
20k
25k
30k
Transaction Amount (AVG)

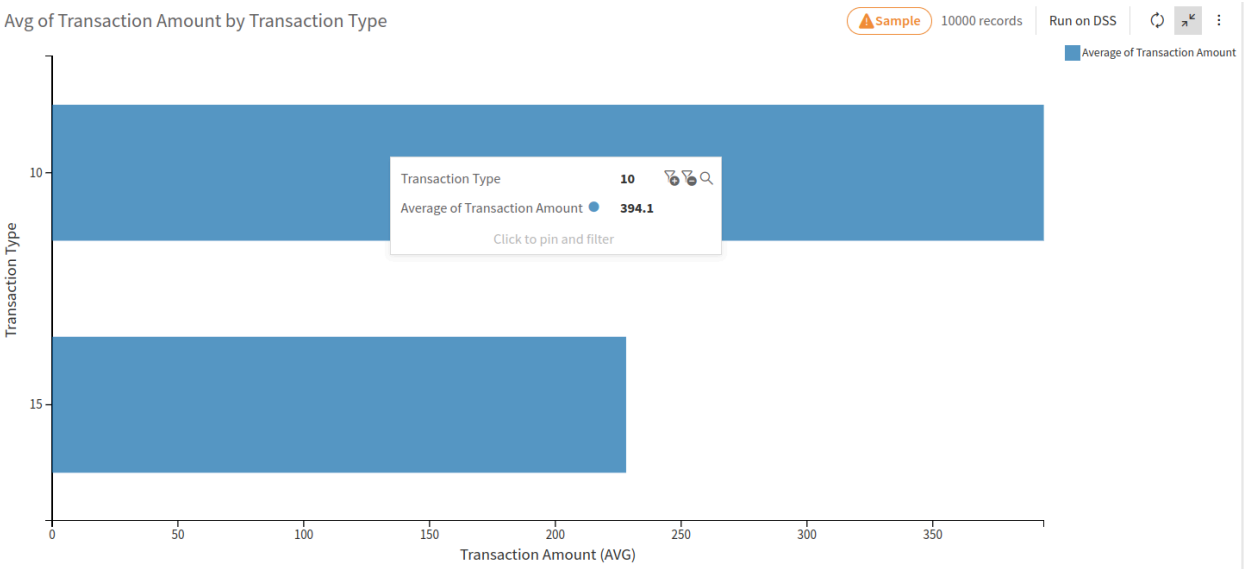

Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
15
Average of Transaction Amount
291
Click to pin and filter
Transaction Type
10
15
0
5000
10k
15k
20k
25k
30k
Transaction Amount (AVG)

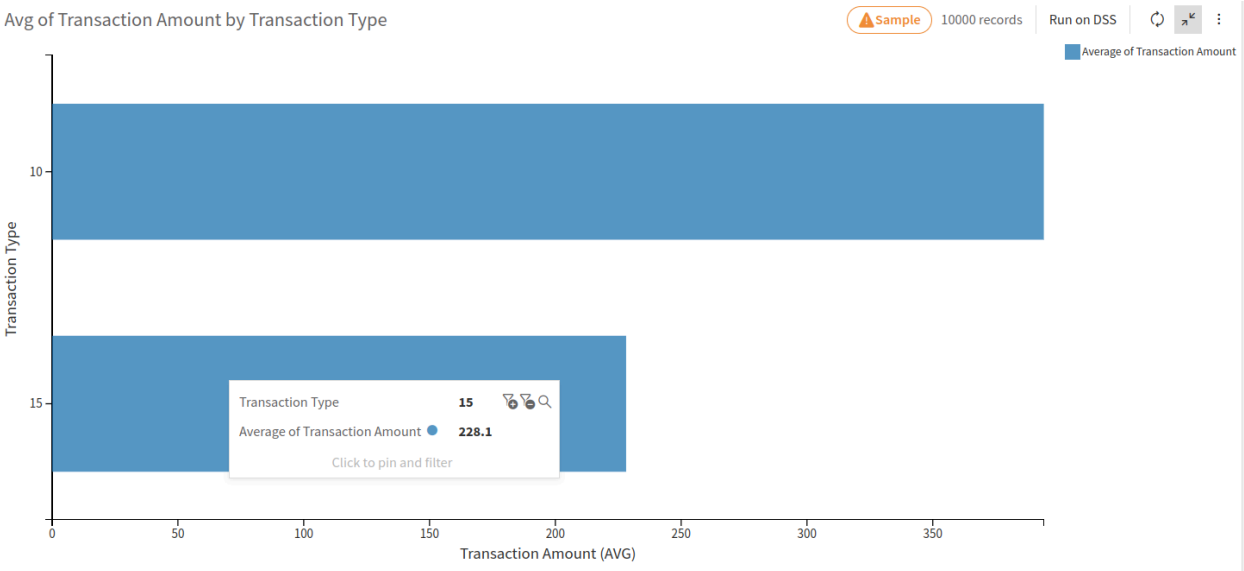
8. 2020_b Data


Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
Average of Transaction Amount
475.8
Click to pin and filter
10
15
Transaction Type
0
50
100
150
200
250
300
350
400
450
Transaction Amount (AVG)

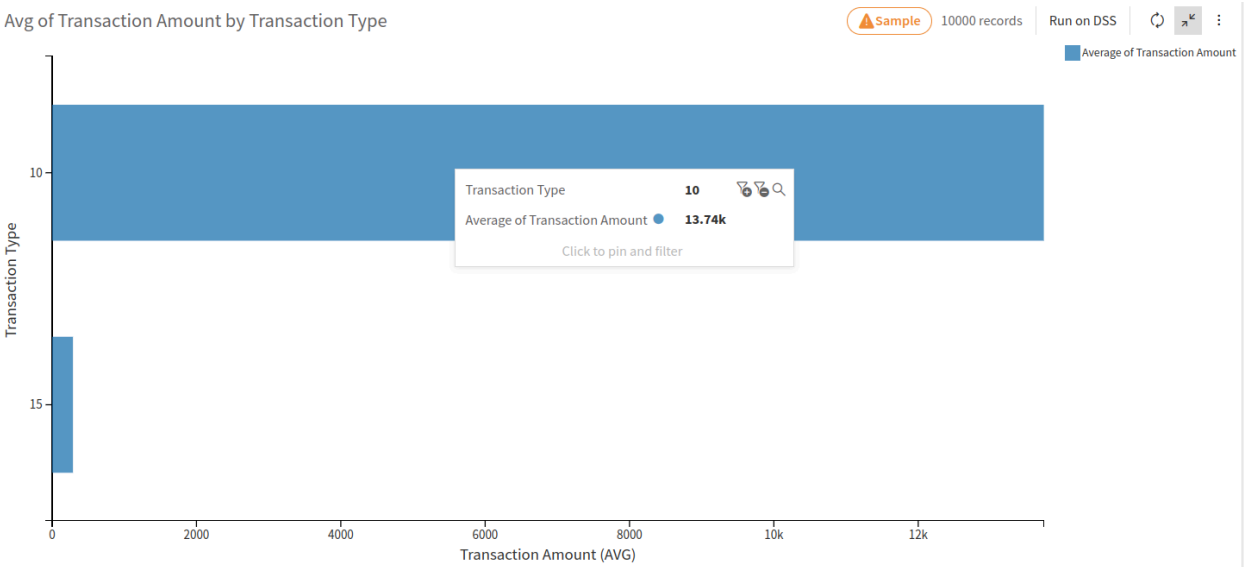

Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
15
Average of Transaction Amount
107.1
Click to pin and filter
10
15
Transaction Type
0
50
100
150
200
250
300
350
400
450
Transaction Amount (AVG)

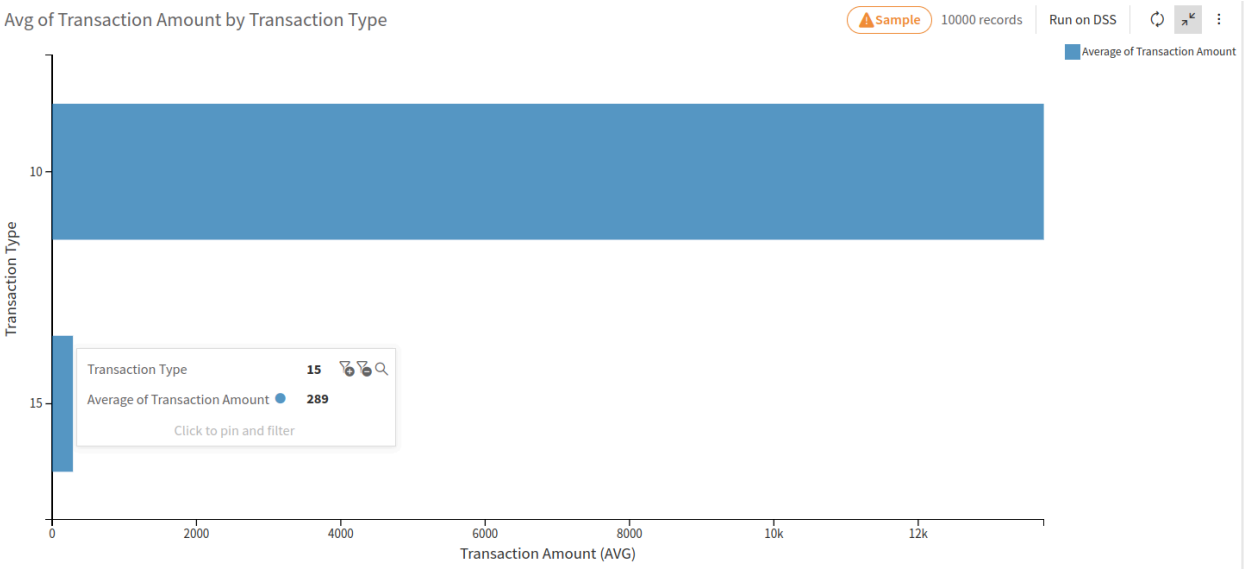
9. 2020_c Data


Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
15
Transaction Type 10
Average of Transaction Amount 394.1
Click to pin and filter
0
50
100
150
200
250
300
350
Transaction Amount (AVG)

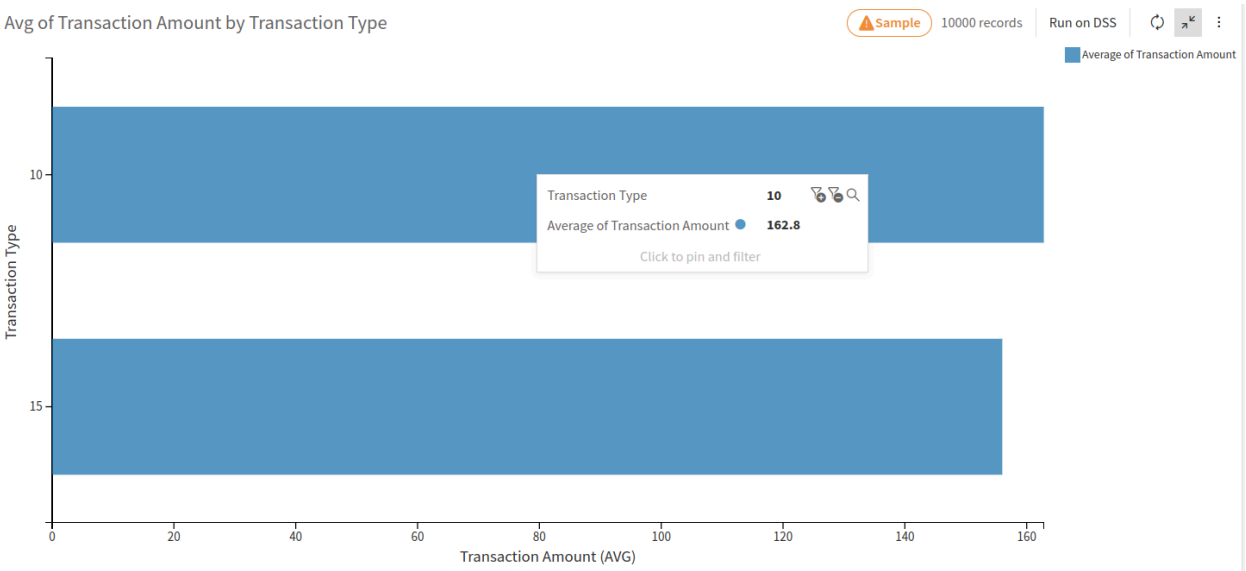

Avg of Transaction Amount by Transaction Type
Sample
10000 records
Run on DSS
Average of Transaction Amount
Transaction Type
10
15
Transaction Type 15
Average of Transaction Amount 228.1
Click to pin and filter
0
50
100
150
200
250
300
350
Transaction Amount (AVG)

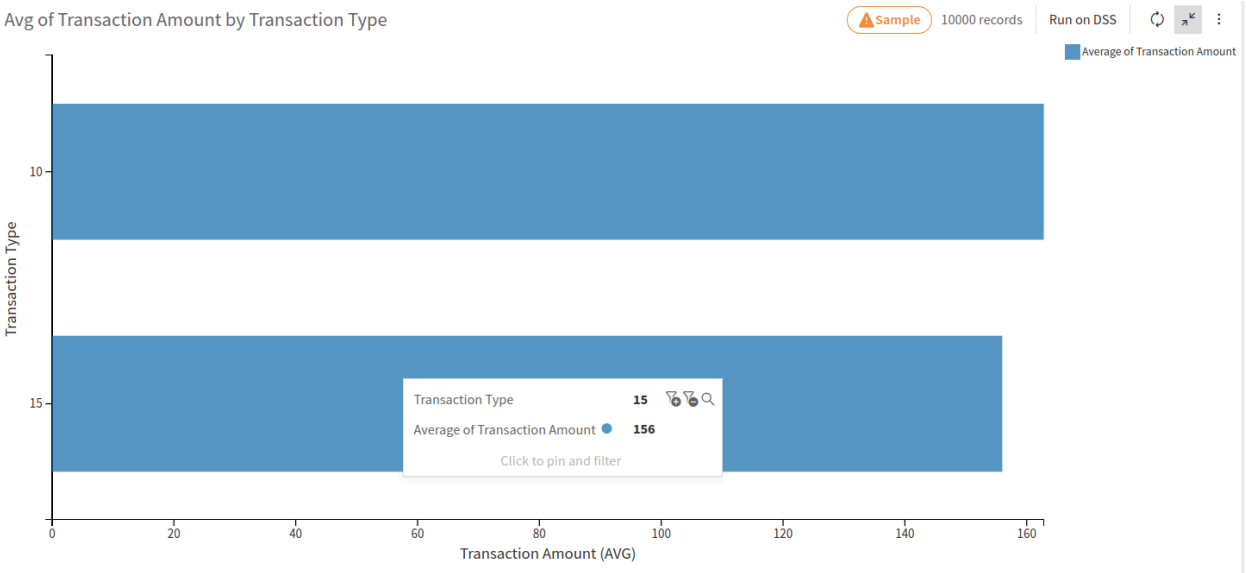
10. 2022_a Data





11. 2022_b Data

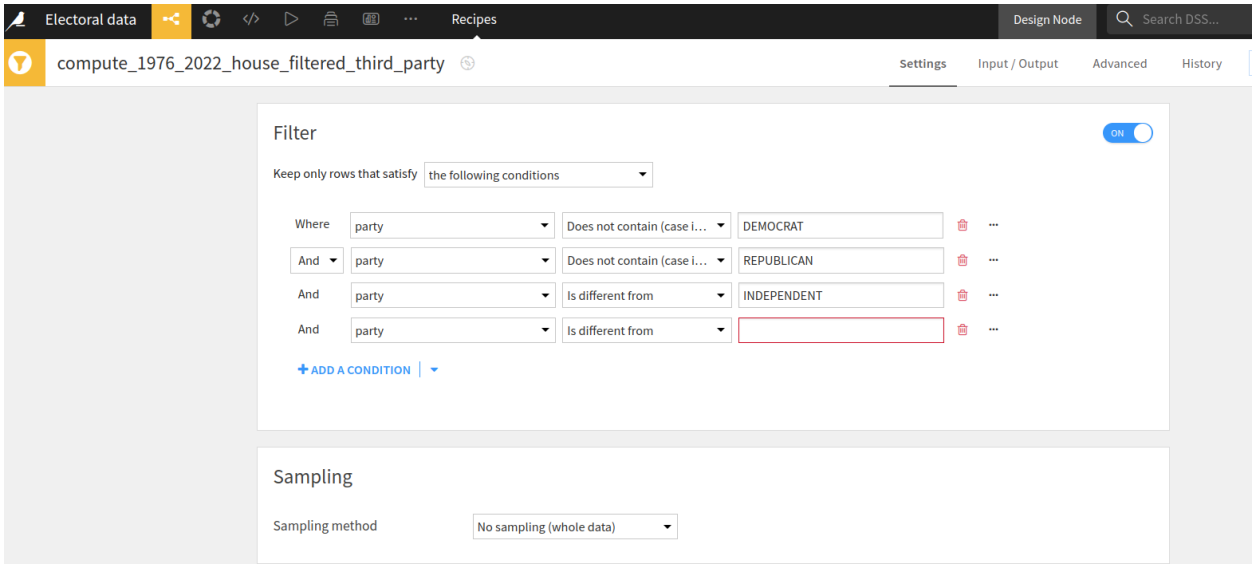




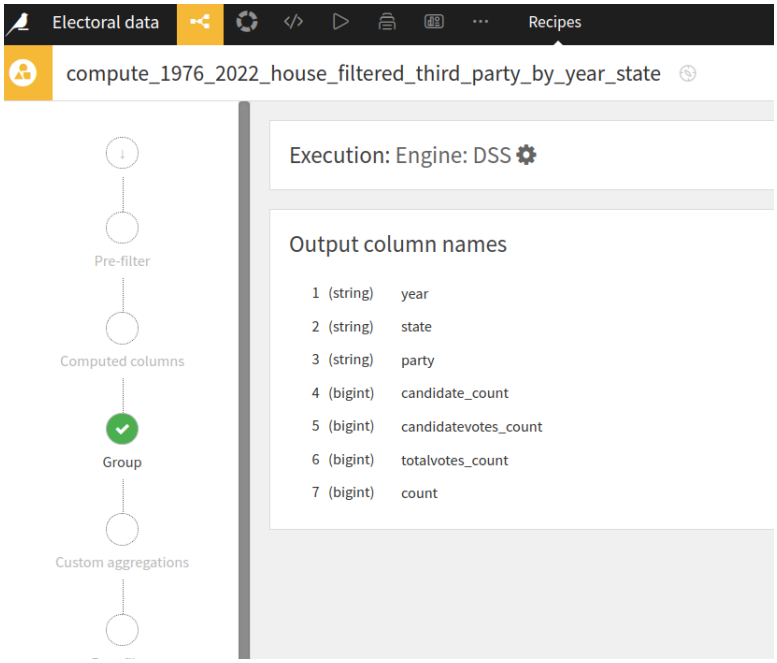
Question for EDA

Do the number of third party candidates differ (avg/state) over time?

Filter the data (Same Operation has been applied to subsequent Datasets.)
We will filter out DEMOCRAT , REPUBLICAN and INDEPENDENT parties from the dataset , so that we have third parties only.
We also remove the Null values with no entries.



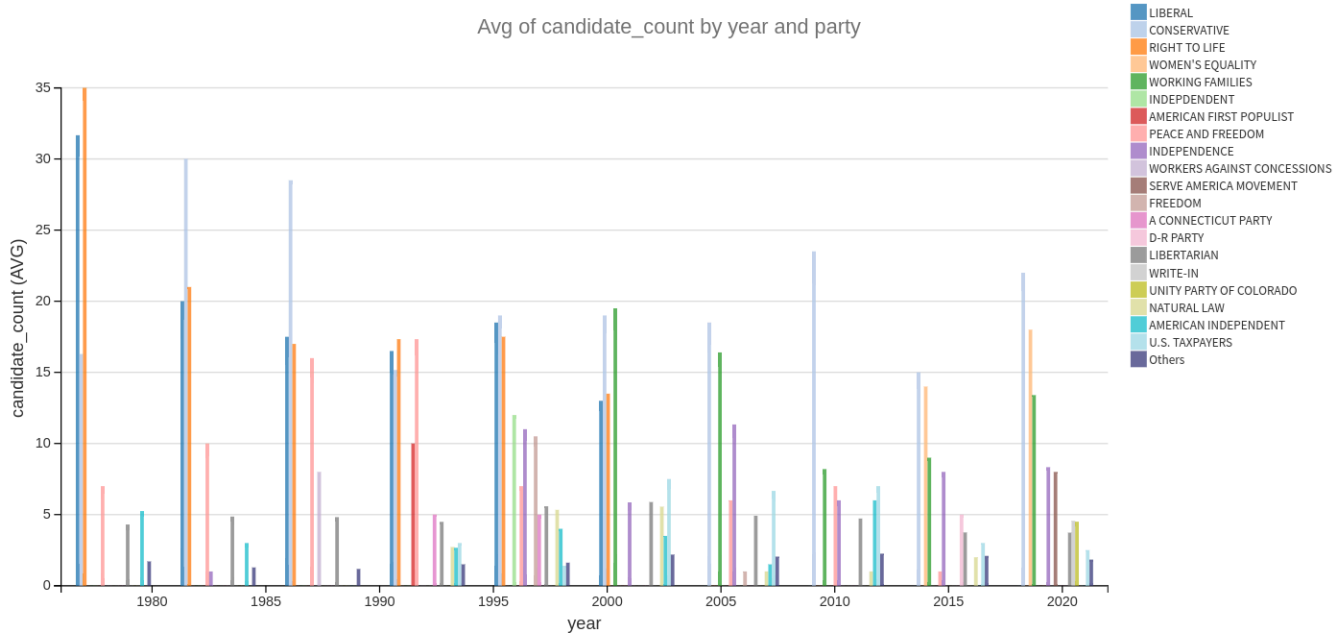
Group by Year and State and Count the number of distinct candidates from each third party candidates.



This will group the sparse data that is present in the first table, to visualize the trend of Third-party candidate participation increase. We have grouped the first two columns Year , State and the party (third-party) , then count the number of candidates that participated during that year in that state.

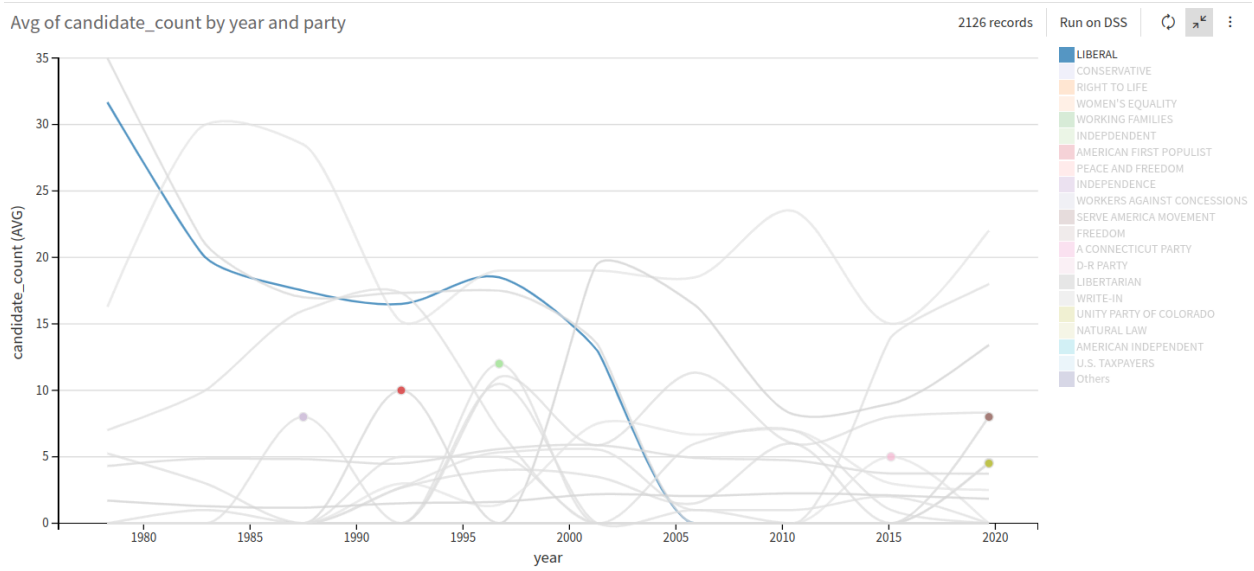
year	state	state_po	state_fips	state_cen	state_ic	office	district	stage	runoff	special	candidate	party
string	string	string	string	string	string	string	string	string	string	string	string	string
Integer	US State	US State	Integer	Integer	Integer	Text	Integer	Text	Boolean	Boolean	Text	Text
1976	ALABAMA	AL	1	63	41	US HOUSE	3	GEN	FALSE	FALSE	OGBURN GARDNER	PROHIBITION
1976	ARIZONA	AZ	4	86	61	US HOUSE	1	GEN	FALSE	FALSE	SUMMER DUNCAN DODGE	LIBERTARIAN
1976	ARIZONA	AZ	4	86	61	US HOUSE	2	GEN	FALSE	FALSE	MICHAEL EMERLING	LIBERTARIAN
1976	ARIZONA	AZ	4	86	61	US HOUSE	4	GEN	FALSE	FALSE	PAT HARPER	LIBERTARIAN

year	state	party	candidate_count	candidatevotes_count	totalvotes_count
string	string	string	bigint	bigint	bigint
Integer	US State	Text	Integer	Integer	Integer
1976	ALABAMA	PROHIBITION	1	1	1
1976	ARIZONA	LIBERTARIAN	3	3	3

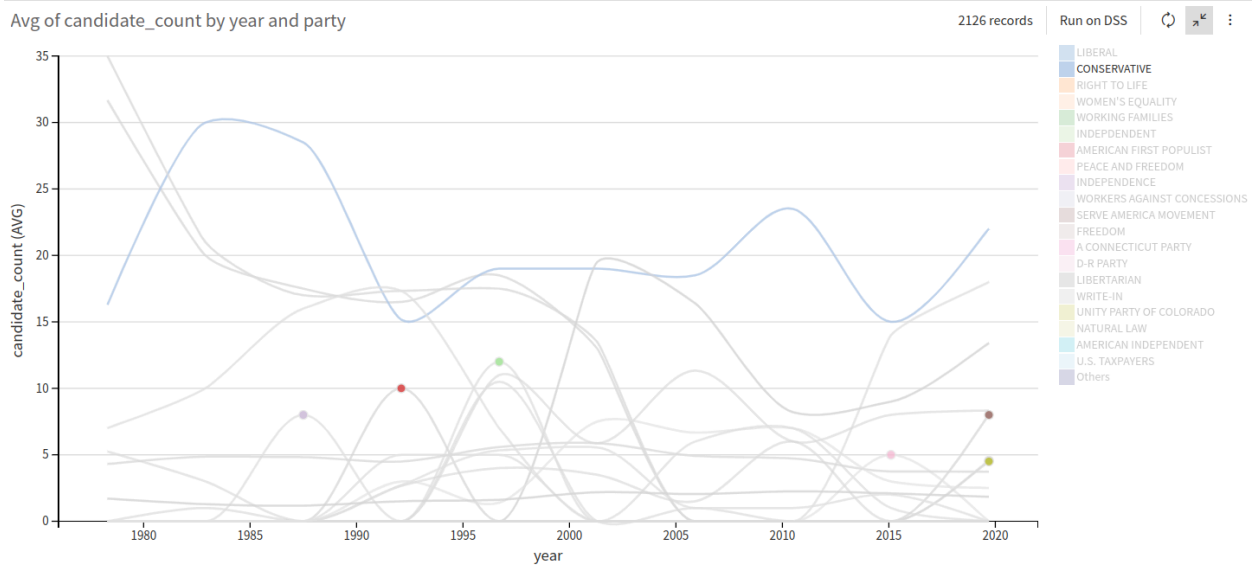


Do the lifespan of third parties differ?
I will post 3 of them here but will store all the screenshot on the github.
EDA/lifespan for all the parties.

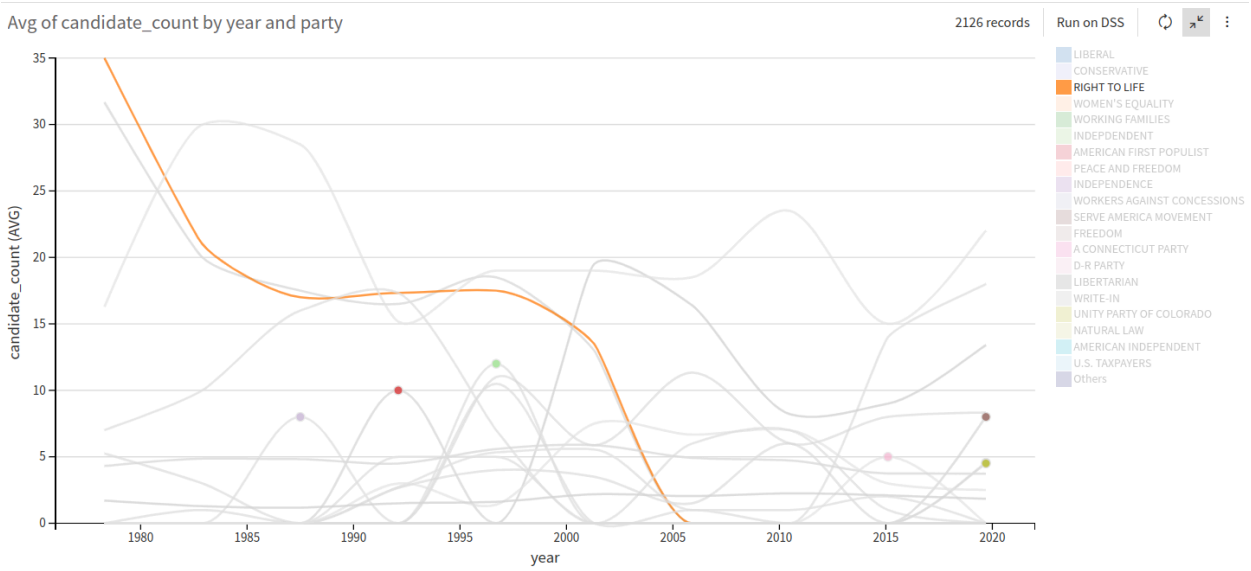
LIBERAL



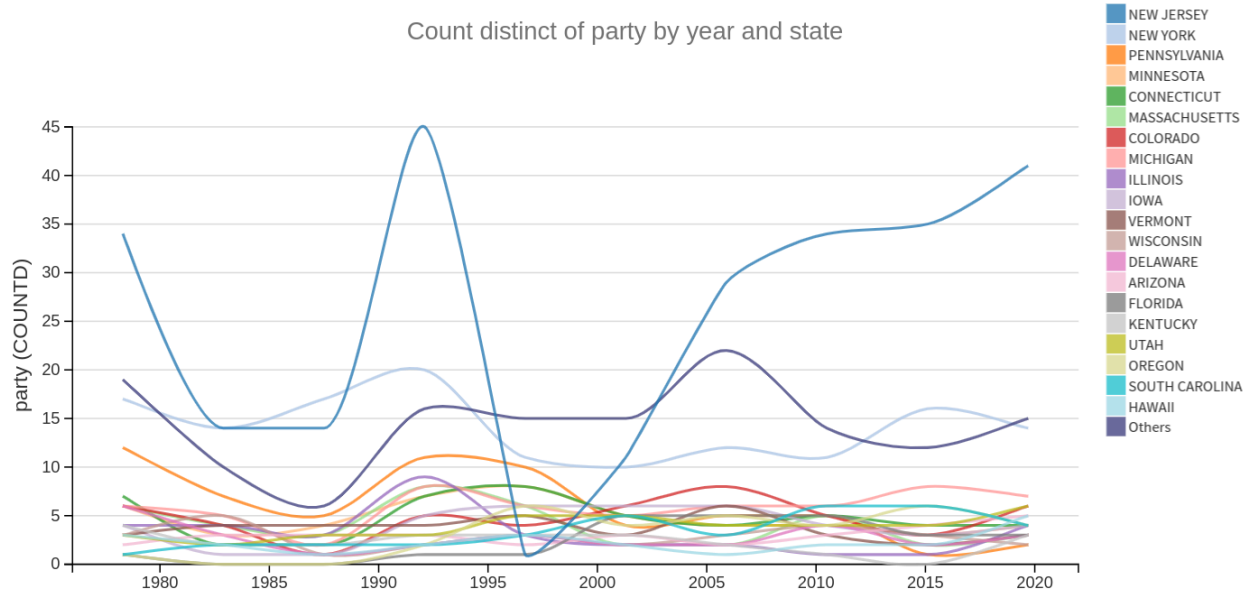
CONSERVATIVE



Right TO LIFE

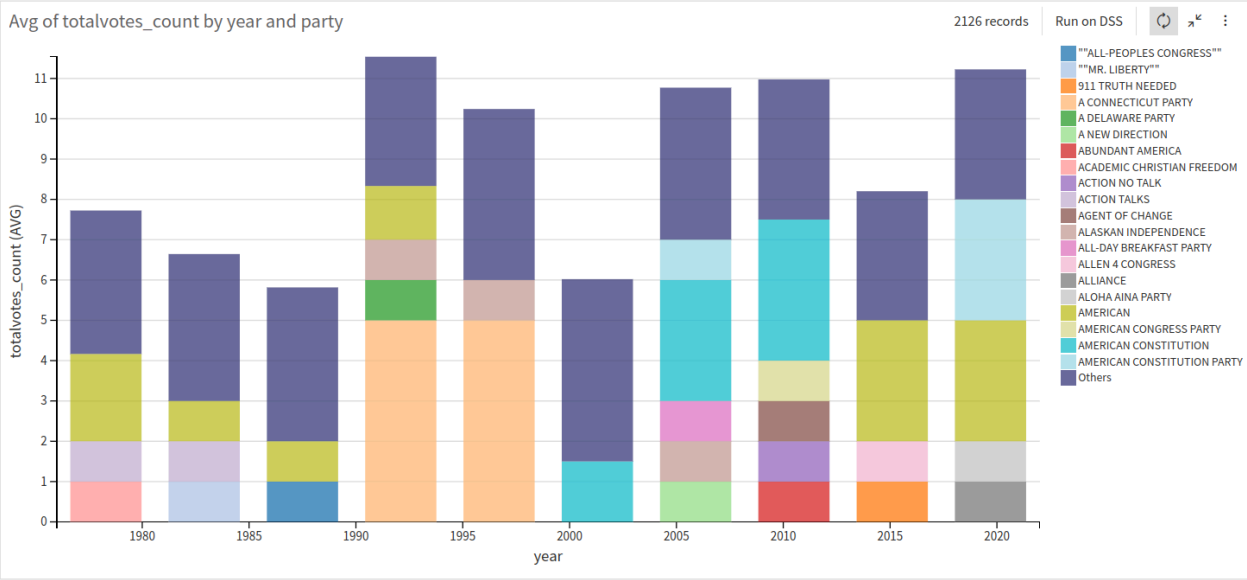


The party number count by State



Do third parties win more elections, year by year?

To analyze the following we had to use the total vote count we aggregated from and visualize it against the year and party name.



Example

Trend for American party