U.S.A Business Formation Statistics.

BY FAROYE OLUMIDE.

Data from States and Counties on business application filings from 2005-2021

About Dataset

Context:

Every year thousands of individuals and entities file business formation applications in the United States. These applications are filed with the state and it is the first step in the process of forming your own business. The United States Census Bureau collects data on these filings and releases a public dataset every quarter.

The business applications measured here are applications for tax IDs from the IRS called Employee Identification Numbers (EIN). An EIN is basically a social security number (SSN) for your business that you must receive from the IRS to open a bank account for your business and complete the application process in your state. So this dataset measures the number of people from each state and county who have filed an application for an Employee Identification Number (EIN).

Content

This dataset was collected from the U.S. Census Bureau. It was part of the annual release of Business Formation Statistics on their website.

This dataset contains two CSV files. One on business application data by State and another on business application data by Country.

Country

State - Abbreviated state name
Country - County that each EIN application came from
Country Code - This is a 4 or 5 digit number that identifies each county. It is made by combining the FIP state and FIP county number.
fipstate - two digit FIPS codes that identify unique geographic areas. Makes up the first two digits in the County Code.
fipcity - A county FIPS code is 2-3 digits that identifies a unique geographic area. It makes up the last 2-3 digits in the County Code.
2005-2019 - Years in which business applications were filed.

State

Year - Year in which EIN application was filed (2006-2021).
Week - The number of the week in the year in which the application was filed (1-52).
State - Abbreviated state name from which EIN application was filed.

Business Applications - This is the total number of EIN applications from that state filed this week.
High-Propensity Applications - This is the number of EIN applications filed that are considered high-propensity applications. Based on the applications filed these businesses have a high propensity (or high likelihood) of transitioning into businesses with a payroll.
Applications w/ Planned Wages - This column consists of the number of high-propensity business applications that indicate a planned date for paying wages.
Applications from Corporations - This column consists of a subset of the high-propensity business applications that came from a corporation or service corporation.

TASK:

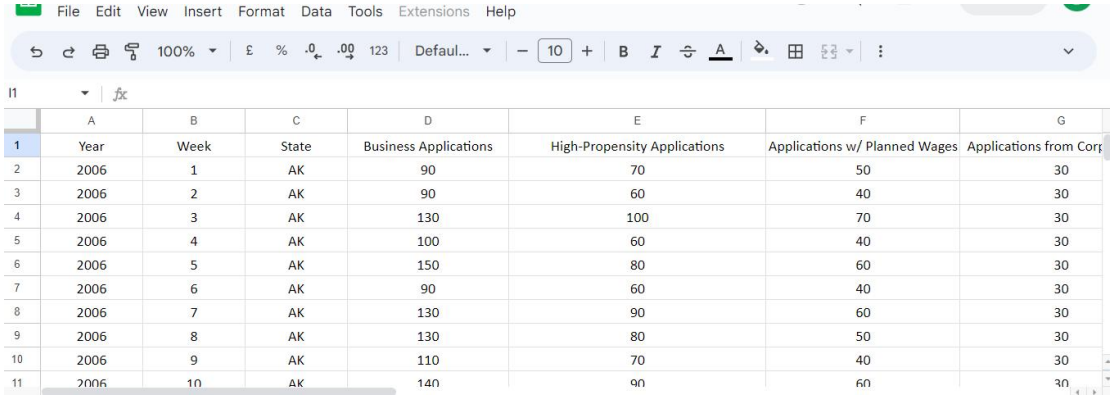
- Analyse the data to determine which states have the highest number of business applications filed?

- Highest percentage that are high-propensity businesses? What is the growth rate year over year?

PROCESS:

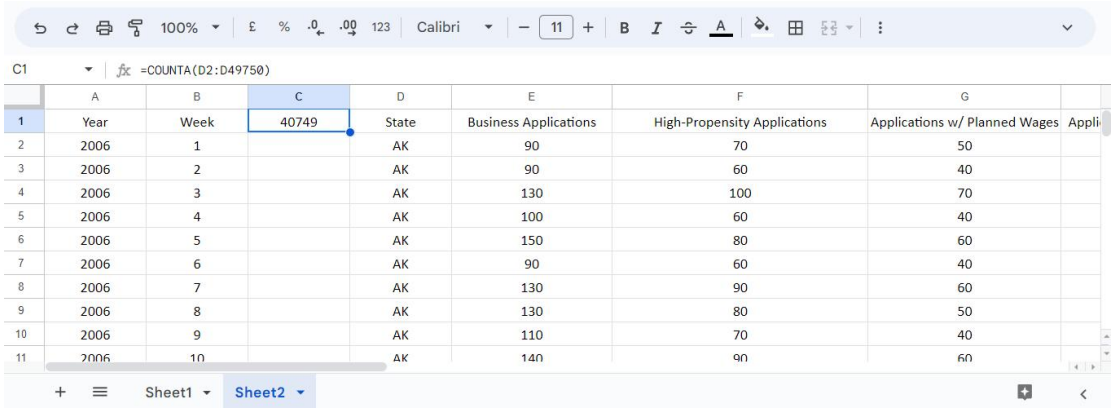
Loading data packages in Goggle sheet for cleaning.

Placing them at the centre



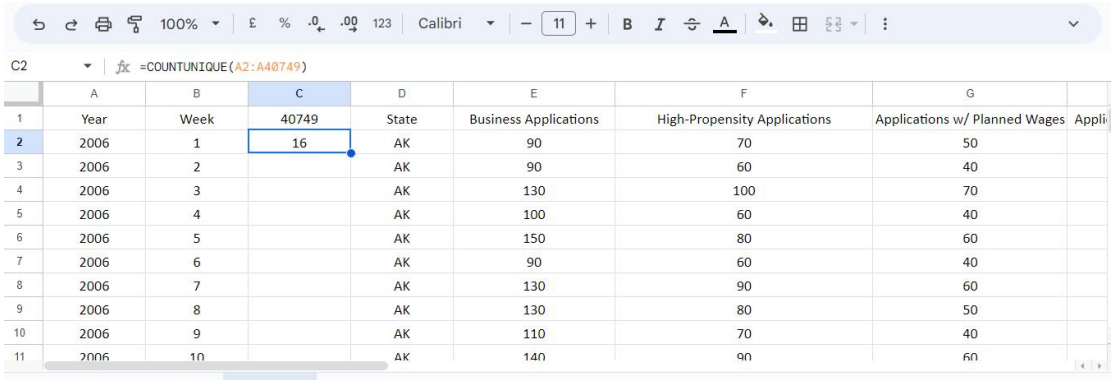
Year	Week	State	Business Applications	High-Propensity Applications	Applications w/ Planned Wages	Applications from Corp
2006	1	AK	90	70	50	30
2006	2	AK	90	60	40	30
2006	3	AK	130	100	70	30
2006	4	AK	100	60	40	30
2006	5	AK	150	80	60	30
2006	6	AK	90	60	40	30
2006	7	AK	130	90	60	30
2006	8	AK	130	80	50	30
2006	9	AK	110	70	40	30
2006	10	AK	140	90	60	30

The COUNTA function was used to determine the total number of States who in the datasets. =COUNTA(D2:D49750) = 40749



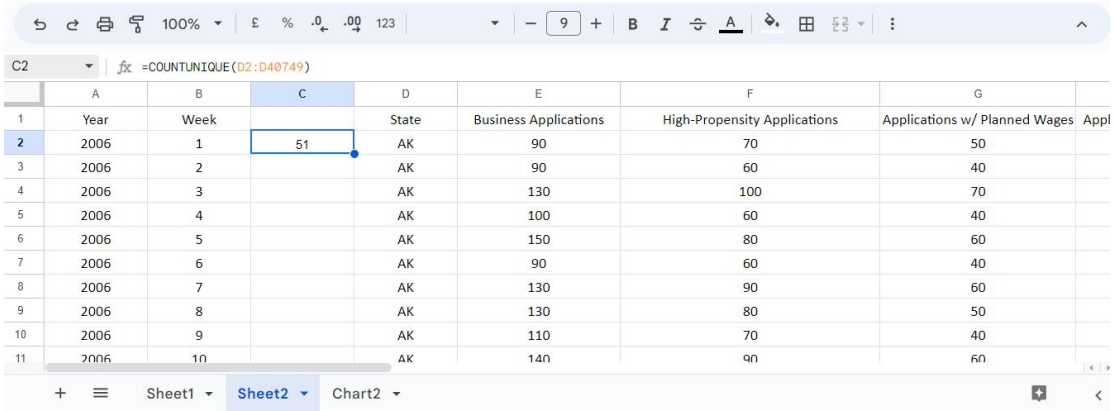
Year	Week	40749	State	Business Applications	High-Propensity Applications	Applications w/ Planned Wages	Appli
2006	1		AK	90	70	50	
2006	2		AK	90	60	40	
2006	3		AK	130	100	70	
2006	4		AK	100	60	40	
2006	5		AK	150	80	60	
2006	6		AK	90	60	40	
2006	7		AK	130	90	60	
2006	8		AK	130	80	50	
2006	9		AK	110	70	40	
2006	10		AK	140	90	60	

The COUNTUNIQUE function was used to determine the total number of Unique years in the dataset where we have 16 different years. =COUNTUNIQUE(A2:A49750) = 16.



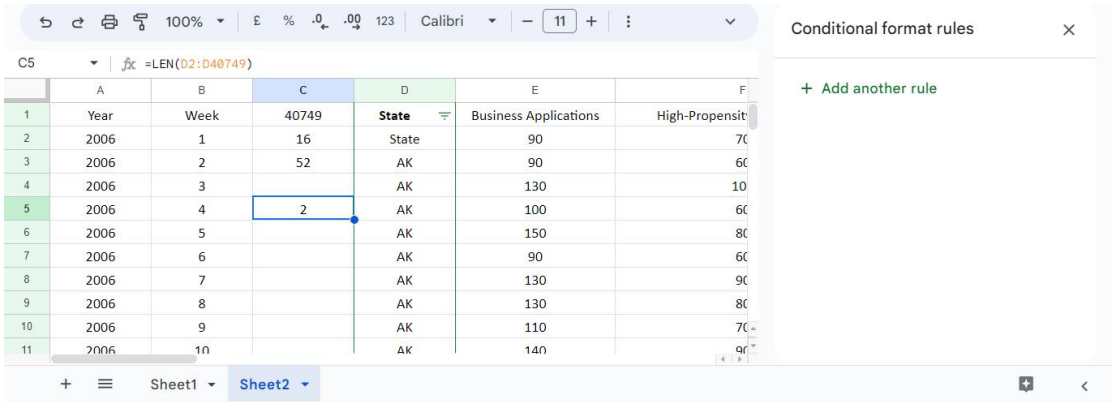
Year	Week	40749	State	Business Applications	High-Propensity Applications	Applications w/ Planned Wages	Appli
2006	1	16	AK	90	70	50	
2006	2		AK	90	60	40	
2006	3		AK	130	100	70	
2006	4		AK	100	60	40	
2006	5		AK	150	80	60	
2006	6		AK	90	60	40	
2006	7		AK	130	90	60	
2006	8		AK	130	80	50	
2006	9		AK	110	70	40	
2006	10		AK	140	90	60	

The COUNTUNIQUE function was used to determine the total number of Unique states in the dataset where we have 51 different states. =COUNTUNIQUE(D2:D49750) = 51.



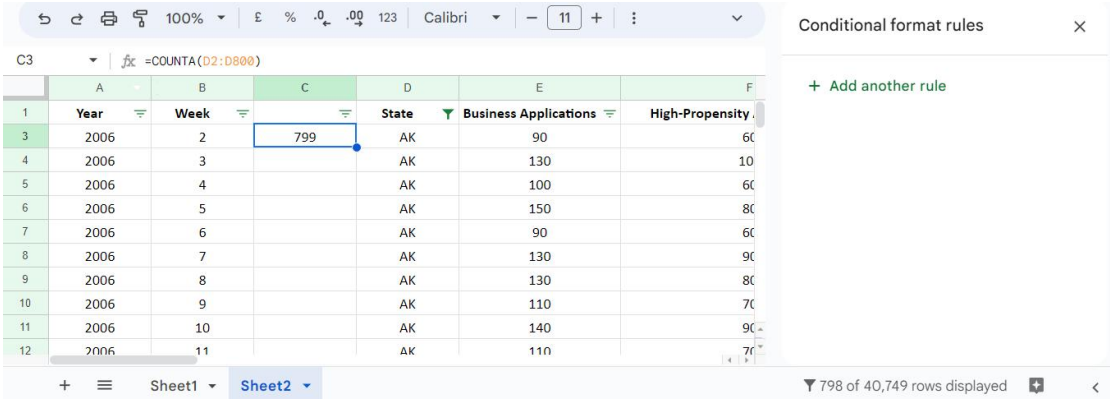
	A	B	C	D	E	F	G
1	Year	Week		State	Business Applications	High-Propensity Applications	Applications w/ Planned Wages
2	2006	1	51	AK	90	70	50
3	2006	2		AK	90	60	40
4	2006	3		AK	130	100	70
5	2006	4		AK	100	60	40
6	2006	5		AK	150	80	60
7	2006	6		AK	90	60	40
8	2006	7		AK	130	90	60
9	2006	8		AK	130	80	50
10	2006	9		AK	110	70	40
11	2006	10		AK	140	90	60

The LEN function was used to determine the length of a string in the state cell which is 2.



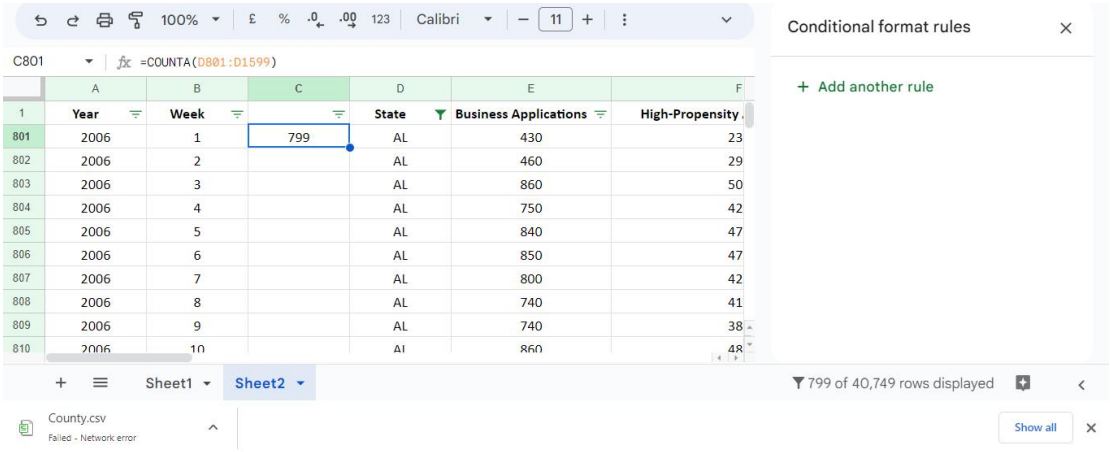
	A	B	C	D	E	F
1	Year	Week	40749	State	Business Applications	High-Propensit
2	2006	1	16	State	90	70
3	2006	2	52	AK	90	60
4	2006	3		AK	130	10
5	2006	4	2	AK	100	60
6	2006	5		AK	150	80
7	2006	6		AK	90	60
8	2006	7		AK	130	90
9	2006	8		AK	130	80
10	2006	9		AK	110	70
11	2006	10		AK	140	90

The COUNTA function was used to determine the total number of times AK State applied for the business application which is 799 times.



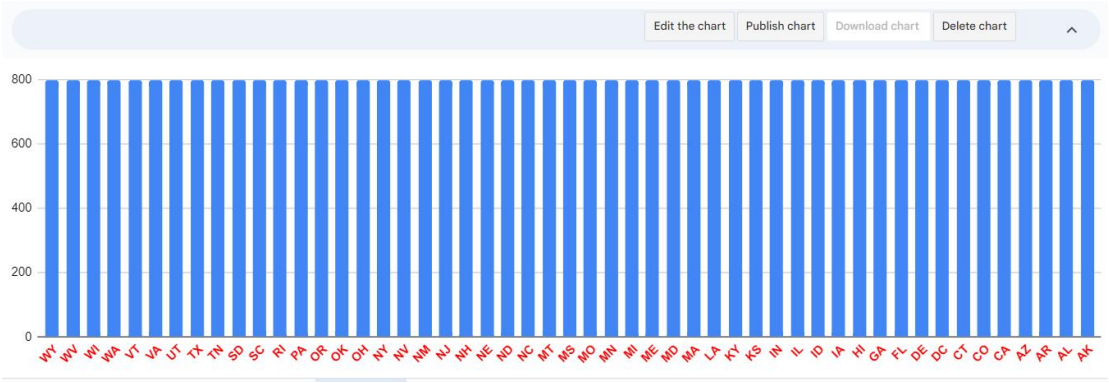
	A	B	C	D	E	F
1	Year	Week		State	Business Applications	High-Propensity
3	2006	2	799	AK	90	60
4	2006	3		AK	130	10
5	2006	4		AK	100	60
6	2006	5		AK	150	80
7	2006	6		AK	90	60
8	2006	7		AK	130	90
9	2006	8		AK	130	80
10	2006	9		AK	110	70
11	2006	10		AK	140	90
12	2006	11		AK	110	70

The COUNTA function was used to determine the total number of times AL State applied for the business application which is 799 times.



	A	B	C	D	E	F
1	Year	Week		State	Business Applications	High-Propensity
801	2006	1	799	AL	430	23
802	2006	2		AL	460	29
803	2006	3		AL	860	50
804	2006	4		AL	750	42
805	2006	5		AL	840	47
806	2006	6		AL	850	47
807	2006	7		AL	800	42
808	2006	8		AL	740	41
809	2006	9		AL	740	38
810	2006	10		AL	860	48

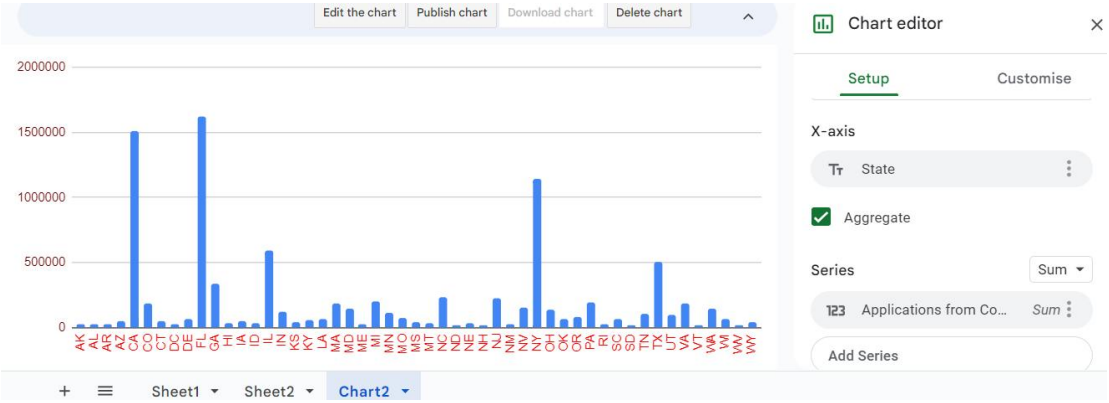
The graph picture below shows all the States that applied for the business application and the result shows that they all applied at 799 times.



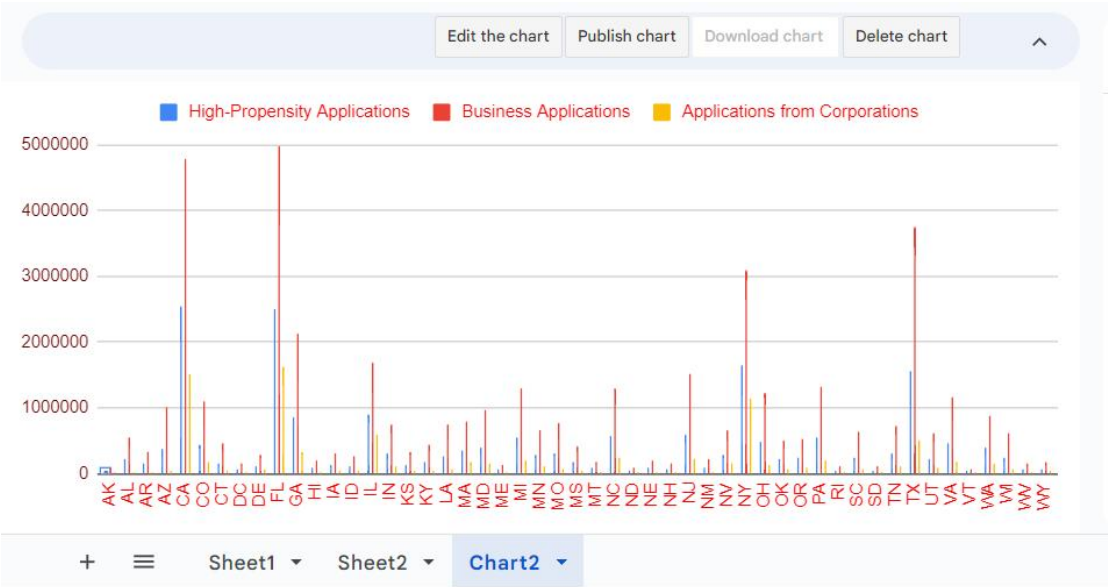
The graph picture below shows all the States that applied for the business applications and the state that have the highest number of business application filed which is FL state.

The graph picture below shows all the States that applied and the state that have the highest number of high-propensity applications which is CA State.

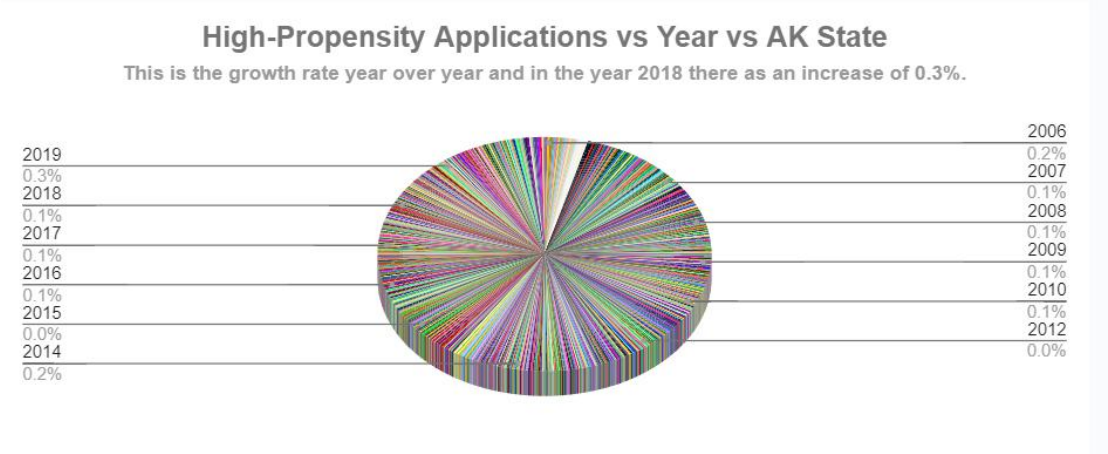
The graph picture below shows all the States that applied from corporate organization and the state that have the highest number is FL State.



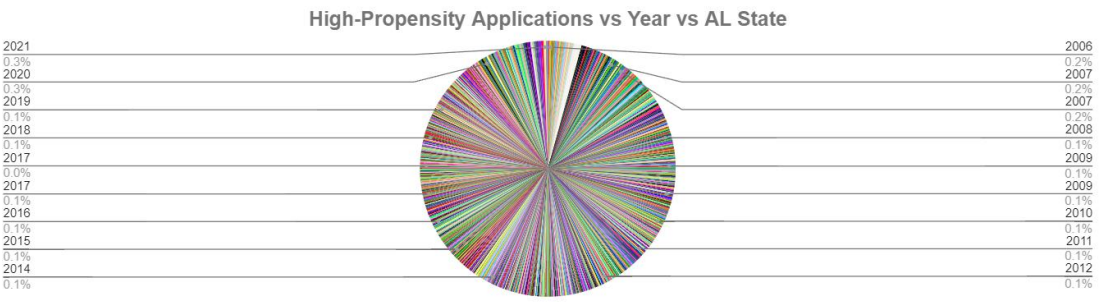
The graph picture below shows all the States that applied and the blue line shows the state with high-propensity application, while the red line shows the business application and the yellow line shows the application from corporations.



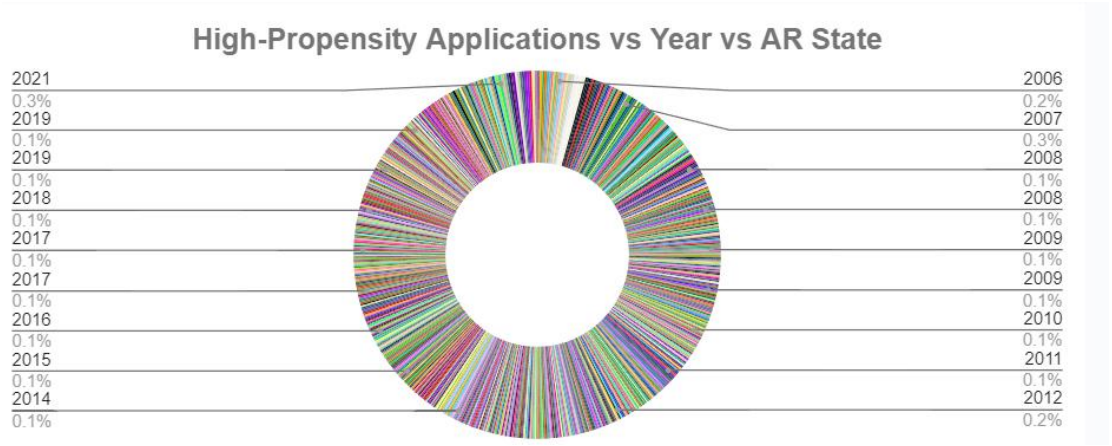
The graph below is the growth rate year over year of AK State and in the year 2019 there as an increase of 0.3% which is the highest in the state.



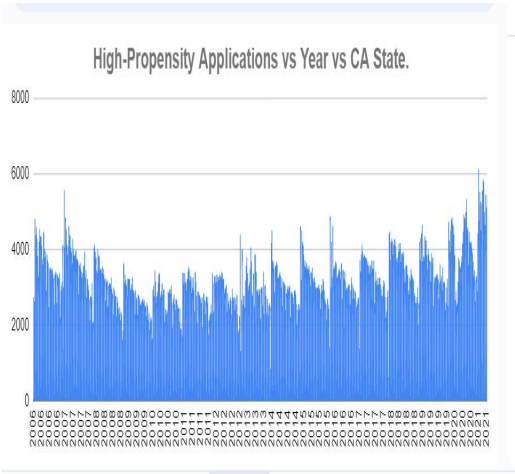
The graph below is the growth rate year over year of AL State and in the year 2020 and 2021 there as an increase of 0.3% which is the highest in the state.

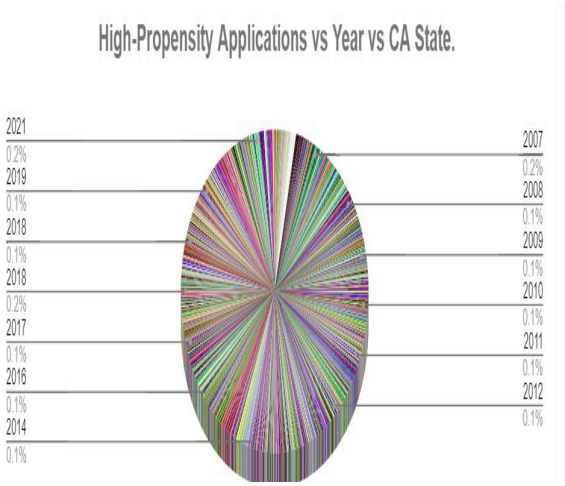


The graph below is the growth rate year over year of AR State and there are increase and fall in between the years .

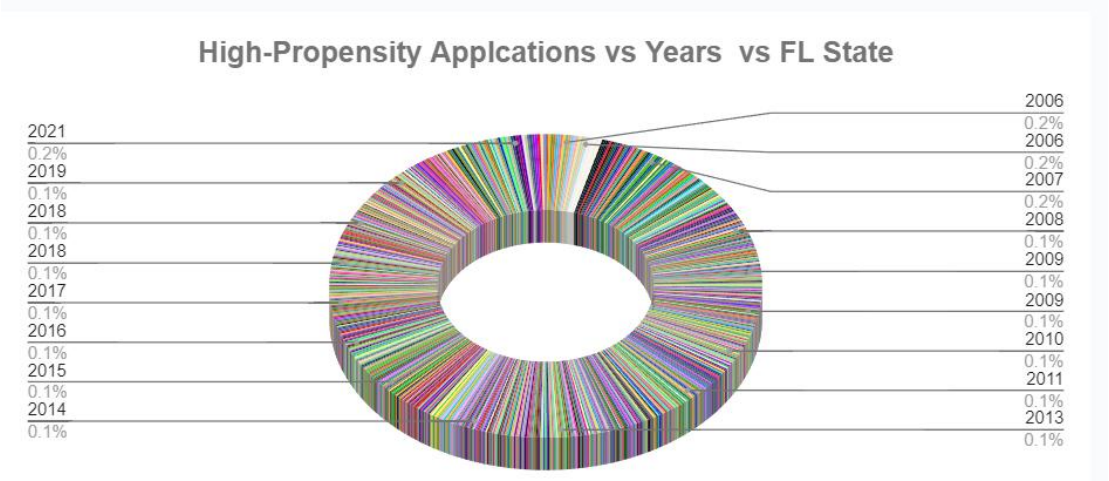


The graph below is the growth rate year over year of CA State and there are increase and fall in between the years .





The graph below is the growth rate year over year of FL State and there are increase and fall in between the years as well .



Summary of key findings.

- After analysing the data the state that have the highest number of business application filed is **FL State**.

- After analysing the data the state that have the highest number of high-propensity applications is **CA State.**

- The growth rate from year over year ranges from 0.1% to 0.3% in different years and increases and decreases in various years as well being shown on the graphs above.

DATA REFENCES:

Data from States and Counties on business application filings from 2005-2021