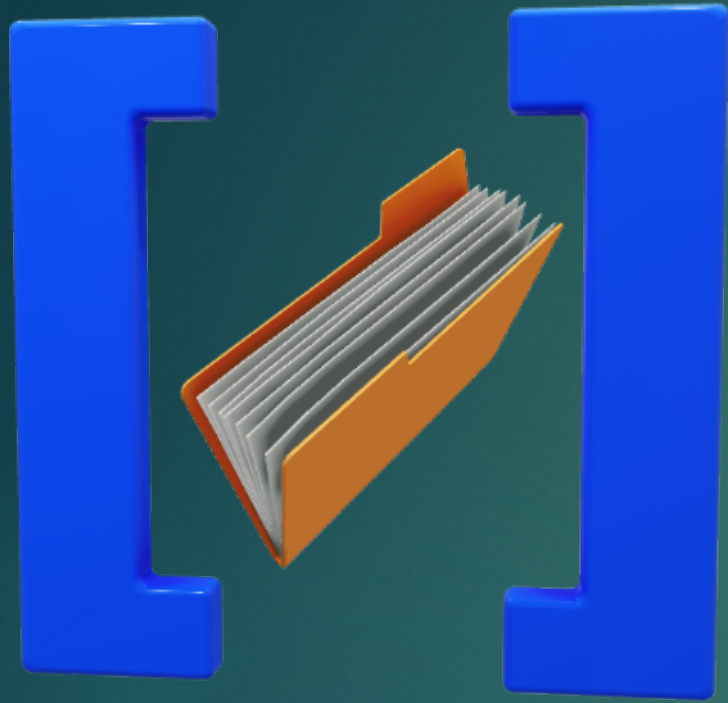




# PROJECT PRESENTATION

ACRA DATA VISUALISATION



# Dataset

- Choice of my Data set will be on [data.gov.sg](https://data.gov.sg) under Acra information entities.
- Base on my selection, I will be choosing this A company data set. I more interesting to find out why Singapore is called the well-known business hub in South East Asia.
- Since the economic is so bad now that I wanted to know whether in Singapore is the company really Sustainable

# Steps by steps

Planning

Solve  
problem  
statement

List problem

Set Question  
to analysis

import all  
project  
csv file

combine  
all csv file  
into a  
dataframe

Create a  
important  
columns

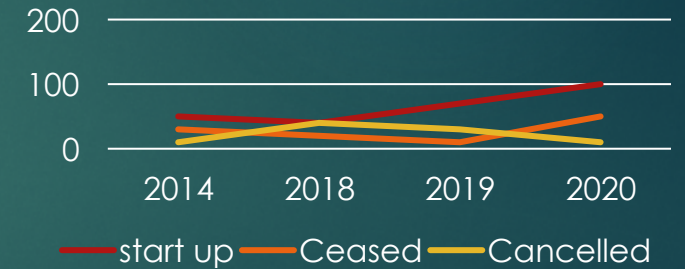
Clean the  
Dataset

Check the  
Dataset if  
its clean

insight

Analyze  
data  
(Graph)

Illustration





# Questions



what is the full address of the company?

Who is the oldest company in Singapore which is on 'live' status?

Is there any company which status is 'Ceased Registration' in Singapore?

Is there any company which unable to sustain in Singapore? > 50years



Is there seasonality for Singapore company which unable to sustain over the years and month?

show the number Company's 'live' status over the years??

show the number Company status 'cancelled' over the years

Over the years, does Singapore company have monthly registration Seasonality?

**the area for business location**

**the popular area for business location**

Over the years, does singapore company able to survive?

# Samples dataset

|   | business_constitution_description | primary_ssic_description                          | primary_user_described_activity                   | street_name           | entity_status_description | annual_return_date | pos |
|---|-----------------------------------|---|---|-----------------------|---------------------------|--------------------|-----|
| 0 | Sole Proprietor                   | REPAIR OF DOMESTIC ELECTRICAL/ELECTRONIC APPLI... | na  | HOUGANG AVENUE 8      | Ceased Registration       | na                 |     |
| 1 | Sole Proprietor                   | CHARTERED BUS SERVICES (INCLUDING SCHOOL BUSES)   | na  | YISHUN STREET 41      | Cancelled (Non-Renewal)   | na                 |     |
| 2 | Sole Proprietor                   | MANAGEMENT CONSULTANCY SERVICES FOR HEALTHCARE... | PROVIDE BUSINESS MANAGEMENT AND CONSULTANCY SE... | BUKIT BATOK STREET 31 | Cancelled                 | na                 |     |
| 3 | Sole Proprietor                   | LETTING AND OPERATING OF SELF-OWNED OR LEASED ... | SALES OF DRINKS & FOODS                           | TANJONG PAGAR PLAZA   | Cancelled                 | na                 |     |
| 4 | Sole Proprietor                   | MANAGEMENT CONSULTANCY SERVICES N.E.C.            | na  | TREVOSE CRESCENT      | Cancelled                 | na                 |     |

| _preference | entity_name                  | paid_up_capital2_others | ... | paid_up_capital10_ordinary | paid_up_capital10_others | uen_of_audit_firm4 | paid_up_capital7_others |
|-------------|------------------------------|-------------------------|-----|----------------------------|--------------------------|--------------------|-------------------------|
| na          | A & E ENGINEERING SERVICES   | na                      | ... | na                         | na                       | na                 | na                      |
| na          | A & Q TRANSIT                | na                      | ... | na                         | na                       | na                 | na                      |
| na          | A AND Y CONSULTANCY SERVICES | na                      | ... | na                         | na                       | na                 | na                      |
| na          | A COFFEE SHOP                | na                      | ... | na                         | na                       | na                 | na                      |
| na          | A DIXIT & ASSOCIATES         | na                      | ... | na                         | na                       | na                 | na                      |

```
1 df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 140209 entries, 0 to 140208  
Data columns (total 95 columns):
```

| #  | Column                            | Non-Null Count  | Dtype   |
|----|-----------------------------------|-----------------|---------|
| 0  | business_constitution_description | 70393 non-null  | object  |
| 1  | primary_ssic_description          | 140209 non-null | object  |
| 2  | primary_user_described_activity   | 50108 non-null  | object  |
| 3  | street_name                       | 139902 non-null | object  |
| 4  | entity_status_description         | 115391 non-null | object  |
| 5  | annual_return_date                | 48983 non-null  | object  |
| 6  | postal_code                       | 139943 non-null | object  |
| 7  | paid_up_capital10_preference      | 0 non-null      | float64 |
| 8  | entity_name                       | 140209 non-null | object  |
| 9  | paid_up_capital2_others           | 435 non-null    | object  |
| 10 | name_of_audit_firm3               | 23 non-null     | object  |
| 11 | address_type                      | 140209 non-null | object  |
| 12 | paid_up_capital9_currency         | 0 non-null      | float64 |
| 13 | paid_up_capital18_ordinary        | 0 non-null      | float64 |
| 14 | paid_up_capital16_others          | 1 non-null      | object  |
| 15 | paid_up_capital16_ordinary        | 1 non-null      | object  |
| 16 | level_no                          | 115294 non-null | object  |
| 17 | paid_up_capital18_others          | 0 non-null      | float64 |

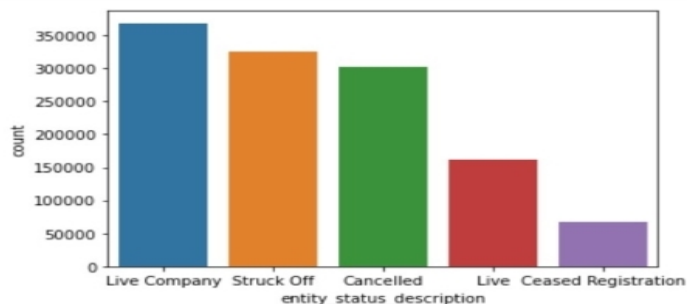
Using a dataset of a single csv file to examine, making sure that all my dataset columns are match before proceeding



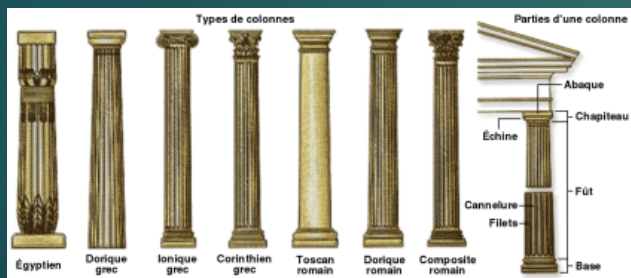
# Insight

## check the insights

```
In [22]: 1 sns.countplot(x="entity_status_description",data = create_df , order = pd.value_counts(create_df['entity_status_description']
2         plt.show())
```



This insight can be seen the highest data among all.



Using the importants and high data to select my columns and check through the data which I needed the most

|    |                                 |        |          |        |
|----|---------------------------------|--------|----------|--------|
| 1  | primary_ssic_description        | 140209 | non-null | object |
| 3  | street_name                     | 139902 | non-null |        |
| 4  | entity_status_description       | 115391 | non-null |        |
| 6  | postal_code                     | 139943 | non-null | object |
| 7  | entity_name                     | 140209 | non-null | object |
| 10 | address_type                    | 140209 | non-null | object |
| 13 | level_no                        | 115294 | non-null | object |
| 20 | uen                             | 140209 | non-null | object |
| 21 | uen_issue_date                  | 140209 | non-null | object |
| 22 | no_of_officers                  | 140209 | non-null | int64  |
| 25 | no_of_charges                   | 140209 | non-null | int64  |
| 32 | block                           | 139698 | non-null | object |
| 47 | registration_incorporation_date | 140209 | non-null | object |
| 52 | primary_ssic_code               | 140209 | non-null | int64  |
| 59 | unit_no                         | 115037 | non-null | object |
| 69 | entity_type_description         | 140209 | non-null | object |

# Challenge and difficulties

## B) import all project csv file (🐼🐼🐼)

```
In [2]: 1 #i created a variable and loop all the csv files
2 list_of_csv_files = r"C:\Users\garyh\Downloads\project_multiple_excel_files\*.csv"
3 for fname in glob.glob(list_of_csv_files):
4     df=pd.read_csv(fname)
5     print(fname)

C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-a.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-b.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-c.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-d.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-e.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-f.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-g.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-h.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-i.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-j.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-k.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-l.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-m.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-n.csv
C:\Users\garyh\Downloads\project_multiple_excel_files\acra-information-on-corporate-entities-o.csv
```

## C) combine all csv file into a dataframe (🐼🐼🐼)

```
In [3]: 1 #after looping, i concat it
2
3 all_files = glob.glob(os.path.join(list_of_csv_files, r"C:\Users\garyh\Downloads\project_multiple_excel_files\*.csv"))
4 concatenated_df = pd.concat((pd.read_csv(f) for f in all_files), ignore_index=True)

<ipython-input-3-8c858802759b>:4: DtypeWarning: Columns (6) have mixed types.Specify dtype option on import or set low_memory=False.
concatenated_df = pd.concat((pd.read_csv(f) for f in all_files), ignore_index=True)

In [4]: 1 #check the information of the dataframe
2 concatenated_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1713722 entries, 0 to 1713721
Data columns (total 94 columns):
#   Column                                Dtype
...
```

Import all csv files into python

Combine all the files

## F8) Slice first 2 digit from existing column 'postal code', create it to a new column (🐼🐼)

```
In [20]: 1 #convert the column to pd Series in order to slice it
2 number = new_df.postal_code.to_dict()
3 s = pd.Series(number)
4 new_df['new_postal_code'] = s.str.slice(stop=2)
5 new_df.head()

Out[20]:
```

| ode | no_of_officers | entity_type_description | new_uen_issue_date | new_registration_date | new_year | new_month | new_year_month | no_of_years | new_postal_code |
|-----|----------------|-------------------------|--------------------|-----------------------|----------|-----------|----------------|-------------|-----------------|
| 446 | 1              | Business                | 2008-09-12         | 2004-01-07 14:33:42   | 2004     | 1         | 2004-01        | 17.0        | 53              |
| 456 | 1              | Business                | 2017-03-15         | 2017-03-15 12:31:24   | 2017     | 3         | 2017-03        | 4.0         | 76              |

Slice out the first 2 columns

```
In [51]: 1 ##find district from URA website and create a csv file to merge it in
2
3 mapping_df = pd.read_csv(r"C:\Users\garyh\Downloads\project_postal_code\postal_code_general_location.csv")
4 mapping_df.head()
5
```

```
Out[51]:
```

| new_postal_code | general_location                              |
|-----------------|---|
| 0               | 1 Raffles Place, Cecil, Marina, People's Park |
| 1               | 2 Raffles Place, Cecil, Marina, People's Park |
| 2               | 3 Raffles Place, Cecil, Marina, People's Park |
| 3               | 4 Raffles Place, Cecil, Marina, People's Park |
| 4               | 5 Raffles Place, Cecil, Marina, People's Park |

19 a) format the column 'new\_postal\_code' from 1 digit to a 2 digit number (🐼)

```
In [52]: 1 mapping_df = pd.read_csv(r"C:\Users\garyh\Downloads\project_postal_code\postal_code_general_location.csv", converters = {'ne
2 mapping_df.head()

Out[52]:
```

| new_postal_code | general_location                               |
|-----------------|--|
| 0               | 01 Raffles Place, Cecil, Marina, People's Park |
| 1               | 02 Raffles Place, Cecil, Marina, People's Park |
| 2               | 03 Raffles Place, Cecil, Marina, People's Park |
| 3               | 04 Raffles Place, Cecil, Marina, People's Park |
| 4               | 05 Raffles Place, Cecil, Marina, People's Park |

Mapping another created csv files and merge to find the region location



| new_postal_code | general_location                                  |
|-----------------|---|
| 53              | Serangoon Garden, Hougang, Punggol                |
| 76              | Yishun, Sembawang                                 |
| 65              | Hillview, Dairy Farm, Bukit Panjang, Choa Chu ... |
| 08              | Anson, Tanjong Pagar                              |
| 29              | Watten Estate, Novena, Thomson                    |

after combination of first 2 digit postal code to find location, results will be shown in the next column in dataframe

```
In [42]: 1 new_df['full_address'] = new_df['block'] + " " + new_df['street_name'] + " " + new_df['level_no'] + " " + new_df['unit_no']
2 new_df.head()
```

```
Out[42]:
```

| _description | new_uen_issue_date | new_registration_date | new_year | new_month | new_year_month | no_of_years | new_postal_code | general_location                                  | full_address                                      |
|--------------|--------------------|-----------------------|----------|-----------|----------------|-------------|-----------------|---|---|
| Business     | 2008-09-12         | 2004-01-07 14:33:42   | 2004     | 1         | 2004-01        | 17.0        | 53              | Serangoon Garden, Hougang, Punggol                | 446 HOUGANG AVENUE 8 B1 1635 530446               |
| Business     | 2017-03-15         | 2017-03-15 12:31:24   | 2017     | 3         | 2017-03        | 4.0         | 76              | Yishun, Sembawang                                 | 455 YISHUN STREET 41 06 75 DEW SPRING @ YISHUN... |
| Business     | 2008-09-13         | 2005-04-29 01:38:25   | 2005     | 4         | 2005-04        | 16.0        | 65              | Hillview, Dairy Farm, Bukit Panjang, Choa Chu ... | 52 BUKIT BATOK STREET 31 09 08 MADEIRA, THE 65... |
| Business     | 2008-09-13         | 2006-04-10 10:36:33   | 2006     | 4         | 2006-04        | 15.0        | 08              | Anson, Tanjong Pagar                              | 1 TANJONG PAGAR PLAZA 01 53 062001                |

Add the details to a new column combination for full address

```
1 for col in new_df.columns:
2     pct_missing = np.mean(new_df[col].isnull())
3     print('{} - {}'.format(col, round(pct_missing*100)))
```

business\_constitution\_description - 0%  
primary\_ssic\_description - 0%  
entity\_status\_description - 0%  
uen - 0%  
entity\_name - 0%  
street\_name - 0%  
block - 0%  
unit\_no - 0%  
level\_no - 0%  
building\_name - 0%  
postal\_code - 0%  
no\_of\_officers - 0%  
entity\_type\_description - 0%  
new\_uen\_issue\_date - 0%  
new\_registration\_date - 0%  
new\_year - 0%  
new\_month - 0%  
new\_year\_month - 0%  
no\_of\_years - 0%  
new\_postal\_code - 0%  
general\_location - 0%

Check the data set to ensure that its cleaned

End of data cleaning



# Use the insights

## 2) Who is the oldest company in Singapore which is on 'live' status?

```
In [27]: 1. live_company = new_df.set_index('entity_status_description') ## to check inside the columns
2. live_company = new_df[new_df['entity_status_description'] == 'Live'] ## Look for this columns company on 'Live'
3. live_company ## total will have 162163 company which is on live in our small dot
```

|    | business_constitutions_description | primary_sic_description                          | entity_status_description | entity_name                    | uen       | street_name        | block | unit_no | building_name | post |
|----|------------------------------------|--|---------------------------|--------------------------------|-----------|--------------------|-------|---------|---------------|------|
| 0  | Sole Proprietor                    | RETAIL SALE OF FOOD N.E.C.                       | Live                      | A & I FOOD SUPPLIER            | 30848000V | TECK WHYE LANE     | 118   | 706     |               |      |
| 12 | Sole Proprietor                    | RETAIL SALE OF COSTUME JEWELLERY                 | Live                      | A & A ACCESSORIES              | 53091902A | CORNERSTONE WALK   | 238   | 890     |               |      |
| 17 | Sole Proprietor                    | GENERAL CONTRACTORS BUILDING CONSTRUCTION INC... | Live                      | A & A AIRCONDITIONING SERVICES | 53060351H | TAMPINES STREET 43 | 438   | 167     |               |      |
| 19 | Sole Proprietor                    | PASSENGER LAND TRANSPORT N.E.C. (BO PRIVATE CO.) | Live                      | A & A ASSOCIATE                | 53194529H | WOODLANDS CIRCLE   | 748   | 919     |               |      |

## 3) Is there any company which status is 'Ceased Registration' in Singapore ?

```
In [30]: 1. ceased_company = new_df.set_index('entity_status_description') ## to check inside the columns
2. ceased_company = new_df[new_df['entity_status_description'] == 'Ceased Registration'] ## Look for this columns company on 'Ceased Registration'
3. ceased_company.head() ## total will have 68256 company which cannot survive(ceased) in singapore
```

|    |                 |   |                     |  |           |                           |     |     |         |  |
|----|-----------------|---|---------------------|--|-----------|---------------------------|-----|-----|---------|--|
| 25 | Sole Proprietor | CONSULTANCY SERVICES (GENERAL)                    | Ceased Registration | BUSINESS NETWORK                         | 52966000B | WOODLANDS DRIVE 14        | 512 | 83  |         |  |
| 27 | Sole Proprietor | CAR WASHING AND RELATED SERVICES                  | Ceased Registration | A & A CAR GROOMING                       | 53019870C | PHOENIX AVENUE            | 71C |     | PHOENIX |  |
| 30 | Sole Proprietor | OTHER CLEANING SERVICES N.E.C. (EG CLEANING OF... | Ceased Registration | A & A CLEANING AND PEST CONTROL SERVICES | 53045392D | BUKIT BATOK WEST AVENUE 4 | 412 | 308 |         |  |
| 51 | Partnership     | ACADEMIC TUTORING SERVICES (EG TUITION CENTRES)   | Ceased Registration | A & A EDUCATION CENTRE                   | 52841947E | PASIR RIS STREET 71       | 750 | 10  |         |  |

5 rows x 21 columns

## 3) Is there any company which status is 'Ceased Registration' in Singapore ?

```
In [30]: 1. ceased_company = new_df.set_index('entity_status_description') ## to check inside the columns
2. ceased_company = new_df[new_df['entity_status_description'] == 'Ceased Registration'] ## Look for this columns company on 'Ceased Registration'
3. ceased_company.head() ## total will have 68256 company which cannot survive(ceased) in singapore
```

|    | business_constitutions_description | primary_sic_description                          | entity_status_description | entity_name                | uen       | street_name        | block | unit_no | building_f |
|----|------------------------------------|--|---------------------------|----------------------------|-----------|--------------------|-------|---------|------------|
| 8  | Sole Proprietor                    | REPAIR OF DOMESTIC ELECTRICAL/ELECTRONIC APPL... | Ceased Registration       | A & E ENGINEERING SERVICES | 53011372D | HOUSHANG AVENUE 8  | 449   | 1635    |            |
| 25 | Sole Proprietor                    | MANAGEMENT CONSULTANCY SERVICES (GENERAL)        | Ceased Registration       | A & A BUSINESS NETWORK     | 52966000B | WOODLANDS DRIVE 14 | 512   | 83      |            |
| 27 | Sole Proprietor                    | CAR WASHING AND RELATED SERVICES                 | Ceased Registration       | A & A CAR GROOMING         | 53019870C | PHOENIX AVENUE     | 71C   |         | PHOENIX    |

## 4) Is there any company which unable to sustain in Singapore? > 50years

```
In [32]: 1. unable_sustain = new_df.set_index('entity_status_description') ## to check inside the columns
2. unable_sustain = new_df[new_df['entity_status_description'] == 'Cancelled']
3. unable_sustain.head()
```

|    | business_constitutions_description | primary_sic_description                           | entity_status_description | entity_name                   | uen        | street_name           | block | unit_no | building_name |
|----|------------------------------------|---|---------------------------|-------------------------------|------------|-----------------------|-------|---------|---------------|
| 2  | Sole Proprietor                    | MANAGEMENT CONSULTANCY SERVICES FOR HEALTHCARE... | Cancelled                 | A AND Y CONSULTANCY SERVICES  | 53043572A  | BUKIT BATOK STREET 31 | 52    | 06      | MADER T       |
| 3  | Sole Proprietor                    | LETTING AND OPERATING OF SELF-OWNED OR LEASED...  | Cancelled                 | A COFFEE SHOP                 | 530955072K | TANJONG PAGAR PLAZA   | 1     | 53      |               |
| 4  | Sole Proprietor                    | MANAGEMENT CONSULTANCY SERVICES N.E.C.            | Cancelled                 | A DIXIT & ASSOCIATES          | 53038864L  | TREVOSE CRESCENT      | 74    |         |               |
| 9  | Sole Proprietor                    | LETTING AND OPERATING OF SELF-OWNED OR LEASED...  | Cancelled                 | A & B EATING HOUSE            | 33317200X  | ORCHARD ROAD          |       | 10      |               |
| 10 | Sole Proprietor                    | RENOVATION CONTRACTORS                            | Cancelled                 | A & B RENOVATION & ENTERPRISE | 52952360X  | JALAN TAMPAK          | 14    |         |               |

5 rows x 21 columns

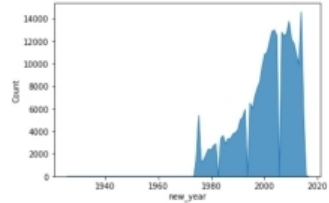
I have created the insights variable in order not to mess up with my data set

# Analyse

9) Is there seasonality for Singapore company which unable to sustain over the years and month?

9a) Over the years

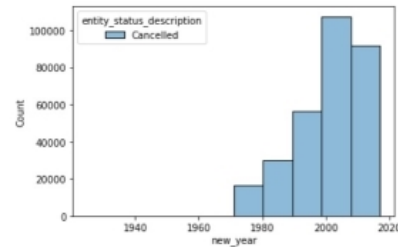
```
In [46]: 1 sns.histplot(data=unable_sustain, x="new_year", element='poly', bins=100)
2 plt.show()
```



9b) monthly over the years

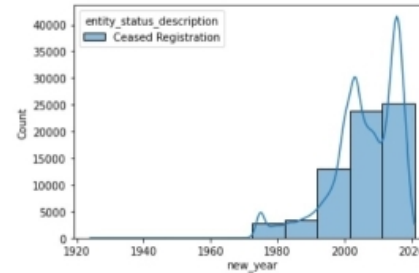
6) show the number Company status 'cancelled' over the years

```
In [36]: 1 sns.histplot(data=unable_sustain, x="new_year", hue="entity_status_description", bins = 10, )
2 plt.show()
```



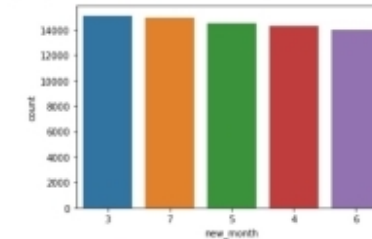
In [38]:

```
1
2 sns.histplot(data=ceased_company, x="new_year", hue="entity_status_description", kde=True, bins=10)
3 plt.show()
```



8) Over the years, does Singapore company have monthly registration Seasonality? ↑

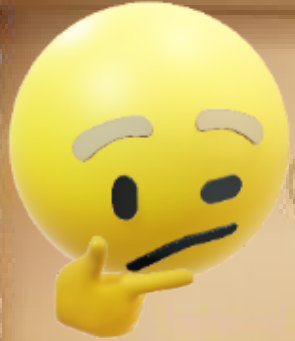
```
In [40]: 1 # multiple="stack"
2 sns.countplot(x="new_month", data = live_company, order = pd.value_counts(live_company['new_month']).iloc[:5].index)
3 plt.show()
```



With the help of those dataframe and created variables

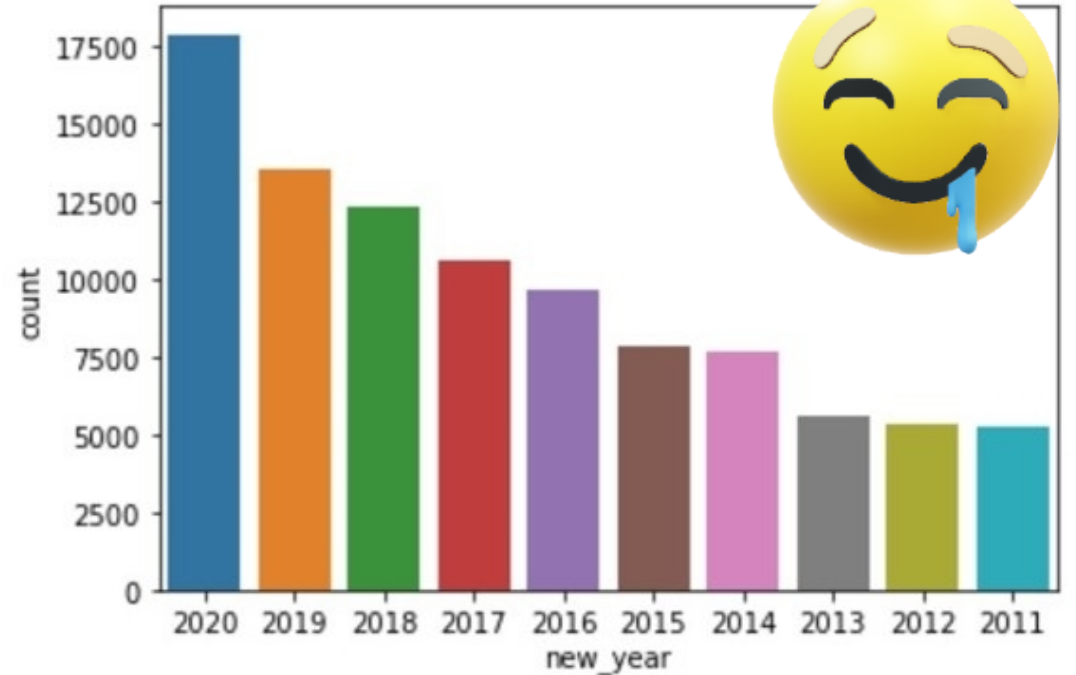


# Problem Statement



## MISSION STATEMENT

Overall Singapore was a well-known as a business hub for southeast Asia, but were the business in Singapore really Serviving??



THE END



<https://github.com/onggreat/G-Dreams>

