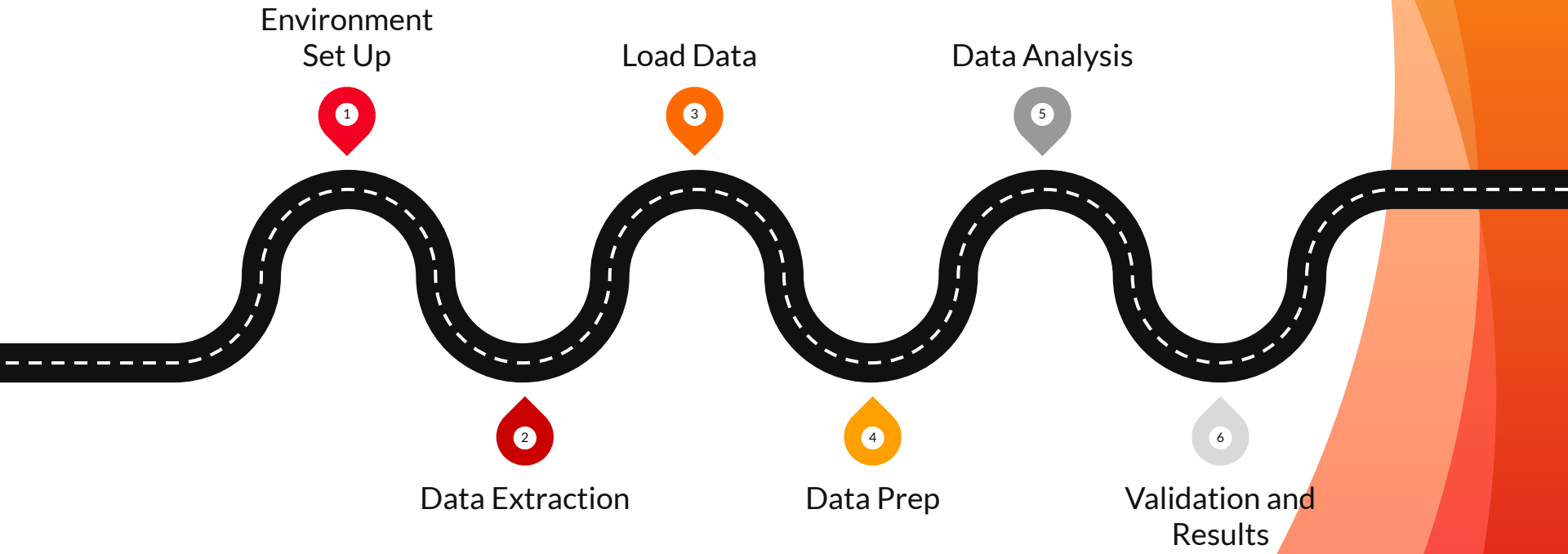
CMS Data Analytics

By Matthew Brenner

Overview


Environment Set Up
1
Load Data
3
Data Analysis
5
2
Data Extraction
4
Data Prep
6
Validation and Results

Environment Set Up

Tech Stack

- Jupyter Notebook on Google Colab cloud environment
- Import Pandas
- Import Numpy

Data Extraction

Extract Data

- Extract most recent LIS Contract Enrollment by County
- Extract most recent Monthly Enrollment by State
- Use !wget Linux command to **extract** data directly from the website[1]
- Use !unzip Linux command to **unzip** file[2]

Healthcare Definitions

Medicare Advantage Plan ("MA" Plans): are offered by Medicare-approved private companies that must follow rules set by Medicare

Low Income Subsidy: a medicare plan that helps people with Medicare pay for prescription drugs, and lowers the costs of Medicare prescription drug coverage.

Data Definitions

State Name - the state name
State Code - the 2 character state code
SSAST - the SSA state code
FIPST - the FIPS (now ANSI) state code
Eligible - the number of Medicare eligibles in the state
MA Enrolled - the number of enrolled MA organizations in the state
Other Enrolled - the number enrolled in other organizations in the state
PDP Enrolled - the number enrolled in PDP organizations in the state

Load Data

Loading Data

- **Read** CSV Files into a Pandas DataFrame[3]
- For LIS enrollment By County, **union** both worksheets to get MA and PDP enrollment[4]

	State Name	State Code	County Name	County Code	Contract#	Plan Name	LIS MA-PD_Enrolled
0	ALABAMA	1	Autauga	1	H0104	BLUE CROSS AND BLUE SHIELD OF ALABAMA	60
1	ALABAMA	1	Autauga	1	H0154	VIVA HEALTH, INC.	655
2	ALABAMA	1	Autauga	1	H0271	UNITEDHEALTHCARE INSURANCE COMPANY OF AMERICA	0
3	ALABAMA	1	Autauga	1	H0432	UNITEDHEALTHCARE OF ALABAMA, INC.	429
4	ALABAMA	1	Autauga	1	H0710	SIERRA HEALTH AND LIFE INSURANCE COMPANY, INC.	0
...	...	...	...	...	...	...	...
77812	VIRGIN ISLANDS	78	St. John/St. Thomas	30	H5774	TRIPLE S ADVANTAGE, INC.	0
77813	VIRGIN ISLANDS	78	St. John/St. Thomas	30	H5991	HEALTH INSURANCE PLAN OF GREATER NEW YORK	0
77814	VIRGIN ISLANDS	78	St. John/St. Thomas	30	H7833	UNITEDHEALTHCARE COMMUNITY PLAN OF TEXAS, L.L.C.	0
77815	VIRGIN ISLANDS	78	St. John/St. Thomas	30	R0759	UNITEDHEALTHCARE INSURANCE COMPANY	0
77816	VIRGIN ISLANDS	78	St. John/St. Thomas	30	R2604	UNITEDHEALTHCARE INSURANCE COMPANY	0

77817 rows × 7 columns

	State Name	State Code	County Name	County Code	Contract#	Plan Name	LIS PDP_Enrolled
0	ALABAMA	1	Baldwin	3	S1030	BCBS OF ALABAMA & UTIC INSURANCE COMPANY	24
1	ALABAMA	1	Baldwin	3	S2668	MEMBERS HEALTH INSURANCE COMPANY	0
2	ALABAMA	1	Baldwin	3	S3285	MG INSURANCE COMPANY	0
3	ALABAMA	1	Baldwin	3	S4802	WELLCARE PRESCRIPTION INSURANCE, INC.	436
4	ALABAMA	1	Baldwin	3	S5552	HUMANA INSURANCE COMPANY OF NEW YORK	0
...	...	...	...	...	...	...	...
42389	VIRGIN ISLANDS	78	St. John/St. Thomas	30	S5820	UNITEDHEALTHCARE INSURANCE COMPANY	53
42390	VIRGIN ISLANDS	78	St. John/St. Thomas	30	S5884	HUMANA INSURANCE COMPANY	0
42391	VIRGIN ISLANDS	78	St. John/St. Thomas	30	S5921	UNITEDHEALTHCARE INS. CO. & UHC INS. CO. OF NY	0
42392	VIRGIN ISLANDS	78	St. John/St. Thomas	30	S5983	MEDCO CONTAINMENT INSURANCE COMPANY OF NEW YORK	0
42393	VIRGIN ISLANDS	78	St. John/St. Thomas	30	S7694	ELIXIR INSURANCE COMPANY	0

42394 rows × 7 columns

Analyze Raw Data

	State Name	State Code	County Name	County Code	Contract#	Plan Name	LIS
0	ALABAMA	1	Autauga	1	H0104	BLUE CROSS AND BLUE SHIELD OF ALABAMA	60
1	ALABAMA	1	Autauga	1	H0154	VIVA HEALTH, INC.	655
2	ALABAMA	1	Autauga	1	H0271	UNITEDHEALTHCARE INSURANCE COMPANY OF AMERICA	*
3	ALABAMA	1	Autauga	1	H0432	UNITEDHEALTHCARE OF ALABAMA, INC.	429
4	ALABAMA	1	Autauga	1	H0710	SIERRA HEALTH AND LIFE INSURANCE COMPANY, INC.	*
...	...	...	...	...	...	...	...
42389	VIRGIN ISLANDS	78	St. John/St. Thomas	30	S5820	UNITEDHEALTHCARE INSURANCE COMPANY	53
42390	VIRGIN ISLANDS	78	St. John/St. Thomas	30	S5884	HUMANA INSURANCE COMPANY	*
42391	VIRGIN ISLANDS	78	St. John/St. Thomas	30	S5921	UNITEDHEALTHCARE INS. CO. & UHC INS. CO. OF NY	*
42392	VIRGIN ISLANDS	78	St. John/St. Thomas	30	S5983	MEDCO CONTAINMENT INSURANCE COMPANY OF NEW YORK	*
42393	VIRGIN ISLANDS	78	St. John/St. Thomas	30	S7694	ELIXIR INSURANCE COMPANY	*

120211 rows × 7 columns

Data Prep

Clean Data

- **Force** all non-numerics to 0 (only for questions 1 and 2)[5]
- **Convert** comma separated numbers to float[6]
- Create a field containing a **unique** value of County and State name

Validation

- Confirm there are no **negative** numbers[7]
- Check **Data Types**[8]

```
State Name      object
State Code       int64
County Name     object
County Code      int64
Contract#       object
Plan Name       object
LIS              int64
dtype: object
```

Data Analysis

Question 1 and 2: Assumptions

- Data points with “*” are equal to 0
- Data points are normally distributed
- Data points may change at source after extraction
- There is no distinction between Medicare Advantage (MA) and Prescription Drug Plan (PDP)
- Medicare members are equal to the amount of people aged 65+

Question 1: LIS Enrollment by State

- **Group** by State[9]
- **Aggregate** by LIS[10]
- **Sort** by LIS in descending order[11]

Question 1: LIS Enrollment by State

State Name	
CALIFORNIA	1621221
NEW YORK	1017347
FLORIDA	982116
TEXAS	890917
PENNSYLVANIA	519515

Question 2: Percent Enrolled

- Create calculated field as **sum** of PDP, MA, and other enrolled[12]
- **Calculate** percentage of enrollments by Total Enrollments/Eligible[14]

Question 2: Percent Enrolled

	STATENAME	STCD	SSAST	FIPSST	Eligible	MA Enrolled	Other Enrolled	PDP Enrolled	Total_Enrolled	Percent_Enrolled
23	Michigan	MI	23	26	2141619	1122535.0	44344.0	984158.0	2151037.0	1.004398
37	Ohio	OH	36	39	2415841	1156268.0	84627.0	961525.0	2202420.0	0.911658
44	South Dakota	SD	43	46	184927	21623.0	30829.0	112635.0	165087.0	0.892714
31	New Jersey	NJ	31	34	1667968	607714.0	1467.0	876229.0	1485410.0	0.890551
18	Kentucky	KY	18	21	954373	440405.0	5357.0	386739.0	832501.0	0.872302
24	Minnesota	MN	24	27	1076421	521207.0	62241.0	349522.0	932970.0	0.866733
40	Pennsylvania	PA	39	42	2820304	1355855.0	12941.0	1067612.0	2436408.0	0.863881

Question 2: Observations

- There may be reporting errors because Michigan has a percentage of over 100%

- “An issue has been identified for the eligibles number, where the number of beneficiaries was double counted for beneficiaries with multiple addresses. The issue has been corrected for the October monthly report.”

Question 2: Validation (October)

STATENAME	Eligible	Total_Enrolled	Percent_Enrolled
Michigan	2137885.0	2133662.0	0.998025
Ohio	2411997.0	2181446.0	0.904415
North Dakota	137826.0	122343.0	0.887663
New Jersey	1661376.0	1471244.0	0.885558
South Dakota	184148.0	161470.0	0.876849

Question 3: Assumptions

- Data points with “*” are treated as missing data (for question 3)
- Data points are normally distributed
- Data points may change at source after extraction

Question 3: Impute Missing Values with Mean

- Fill Missing Values with **mean** of each county[13]
- **Group** county to retrieve estimated LIS count

Question 3: Impute Missing Values with Median

- Fill Missing Values with **median** of each county[14]
- **Group** county to retrieve estimated LIS count

Question 3: Impute Missing Values with Sample

- Calculate **mean** and **standard** deviation of each county

- Use the standard deviation and mean of each county to draw Random sample from a **Gaussian distribution**[15]

Question 3: Measure Uncertainty

- Create **calculated field** of count of values that equal "*" [16]
- **Calculate** percentage of these null values by county/state

	State Name	County Name	LIS	NA_Count	Numeric	NA_Percent	Unique
14180	KENTUCKY	Leslie	70	4	20	16.666667	KENTUCKYLeslie

Question 3: Make Prediction

Unique	State Code	County Code	LIS	LIS_median	LIS_mean	LIS_dis
KENTUCKYLeslie	504	3144	1718.0	1926.0	2061.6	2107.384117

Thanks!

Any questions?

Code Reference

1.
```
!wget -q https://www.cms.gov/files/zip/2021-low-income-subsidy-contract-enrollment-county.zip
!wget -q https://www.cms.gov/files/zip/monthly-enrollment-state-january-2022.zip
```

2.
```
!unzip  2021-low-income-subsidy-contract-enrollment-county.zip
!unzip  monthly-enrollment-state-january-2022.zip
```

3.
```
LIS_file = "/content/2021 LIS PDP&MAPD_by_State_County_Contract.xlsx"
LIS_by_County = pd.read_excel(LIS_file)
```

```
MonthlyEnrollment_file = "/content/Monthly_Report_by_State_2022_01/Monthly_Report_By_State_2022_01.csv
MonthlyEnrollment = pd.read_csv(MonthlyEnrollment_file)
```

4.
```
df = pd.concat([df1,df2])
```

Code Reference

5.
```
df["LIS"] = pd.to_numeric(df["LIS"] , errors ='coerce').fillna(0).astype('int')
```

6.
```
def convertNumCommas(num):
    if type(num) is str:
        if num == "*":
          return 0
        else:
          return int(num.replace(',', ''))
    elif type(num) is int:
        return num
    else:
      return "help"
```

Code Reference

7.
```
df[df["LIS"] <0]
```
8.
```
print(df1.count() + df2.count() == df.count())
```
9.
10.
```
grouped_df = df.groupby('State Name').agg('sum').sort_values("LIS", ascending = False)
```
11.
12.
```
MonthlyEnrollment["Total_Enrolled"] = MonthlyEnrollment["Other Enrolled"]  + MonthlyEnrollment["MA Enrolled"] + MonthlyEnrollment["PDP Enrolled"]
```
13.
```
df_impute = df.replace("*", np.nan)
df_impute["LIS"] = pd.to_numeric(df_impute.LIS, errors = 'coerce')
df_impute["LIS"] = df_impute.groupby("County Name")["LIS"].transform(lambda x: x.fillna(x.mean()))
```
14.
```
df_impute2["LIS"] = df_impute1.groupby("Unique")["LIS"].transform(lambda x: x.fillna(x.median()))
```

Code Reference

15.
```
df_impute2["LIS"] = df_impute1.groupby("Unique")["LIS"].transform(lambda x: x.fillna(np.random.normal(x.mean(), x.std(), 1)[0]))
```

16.

```
df_impute1["NA_Count"] = df_impute1.LIS.isnull().groupby(df_impute1['Unique']).transform('sum').astype(int)
```