Lecture on "Analyze population life expectancy based on Big Data "

At 6pm to 7:30pm on January 7, 2021, the lecture of "Analyze population life expectancy based on Big Data" of School of Silicon Valley Artisan of USJ broadcasted live online globally. The life expectancy of the population is a very important indicator that can comprehensively reflect the level of socio-economic development and the level of medical and health services in a region. The lecture of Silicon Valley Artisan analyzes the factors that affect the life expectancy of the population and tries to use Python 3.7 to build a model to predict the life expectancy of the population and explore its changing trends. The following are the main points summarized by Yuting. Please take notes.

Jupyter 相关性project (autosaved)

File Edit View Insert Cell Kernel Widgets Help
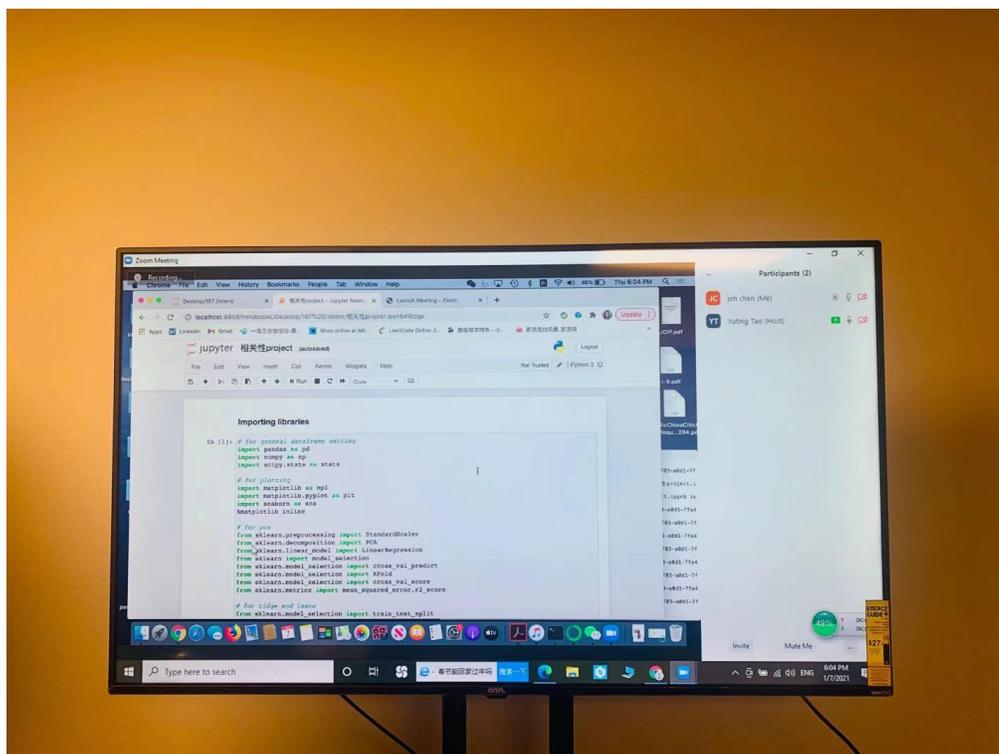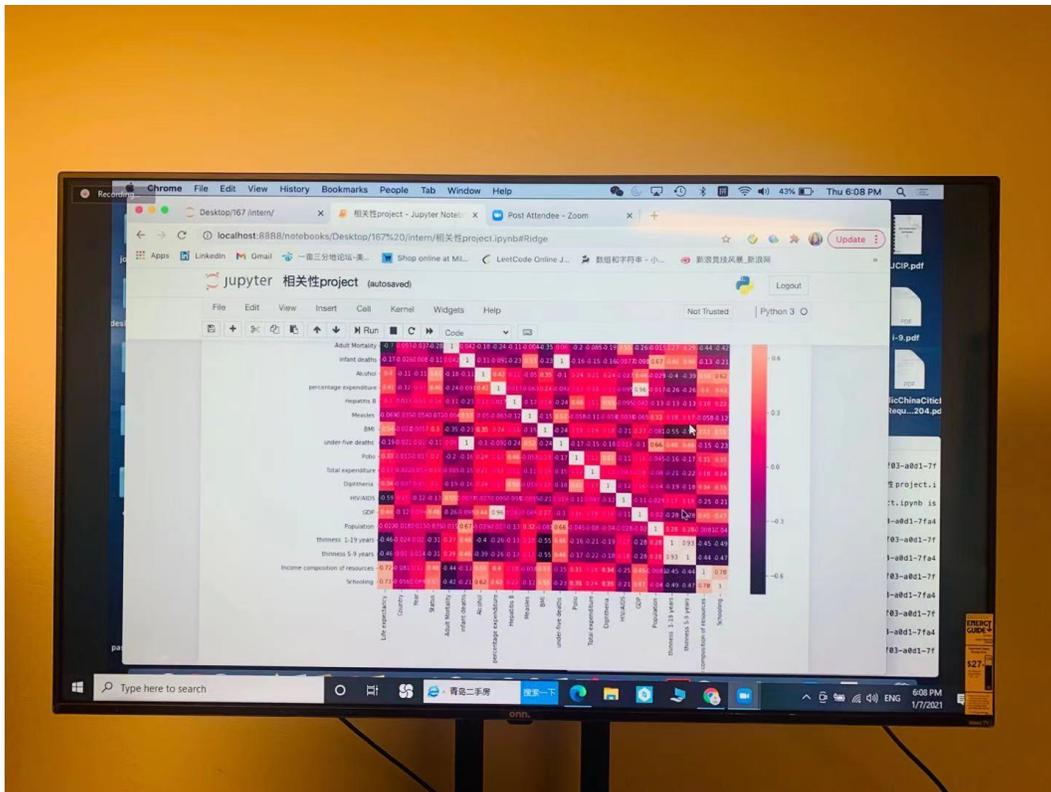
```
explained_variance = pca.explained_variance_ratio_
cumulative_explained_variance=np.cumsum(np.round(explained_variance, decimals=4))
pd.Series(data=cumulative_explained_variance, index=range(1,20))
```

```
Out[14]:  1    0.3300
          2    0.5085
          3    0.6271
          4    0.7103
          5    0.7689
          6    0.8151
          7    0.8506
          8    0.8790
          9    0.9064
          10   0.9283
          11   0.9463
          12   0.9634
          13   0.9769
          14   0.9876
          15   0.9926
          16   0.9956
          17   0.9979
          18   0.9999
          19   1.0000
          dtype: float64
```

```
In [15]: # cumulative variance plot
         plt.rcParams['figure.figsize'] = [8, 5]
         plt.plot(range(1,20),cumulative_explained_variance)
         plt.xticks(range(1,20))
```

I. The review of lecture of "Analyze population life expectancy based on Big Data"

① Project background introduction

Many countries regard the life expectancy of the population as an important index of expectancy, while increasing life expectancy as an important development goal. This project analyzes the factors that affect the life expectancy of the population and determines the degree of impact of each influencing factor on the life expectancy of the population in order to find specific ways to increase the life expectancy of the population. At the same time, the population life expectancy is predicted, and its trend of change is explored.

② Data cleaning

    a. Treatment of missing values and outliers, as well as corrections for obviously wrong data.
    b. Research the characteristics of data variables

③ Model selection

    a. Principal components regression (PCR)
    b. Ridge regression, and least
    c. Absolute shrinkage and selection operator (lasso) regression

④ Model comparison

    a. Coefficient of determination
    b. Mean square error
    c. Model simplicity
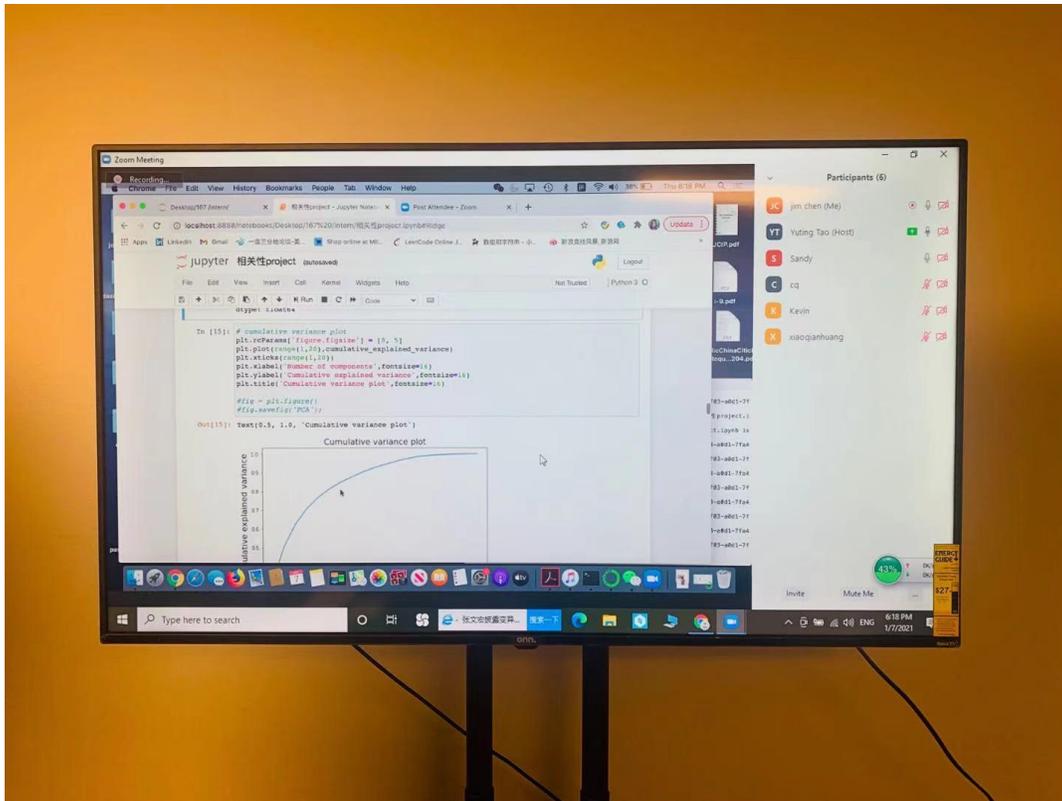
II. Lecture Content

1. Population data with import and clean up
2. Correlation check
3. Select the corresponding model according to the data characteristics
4. Lecture summary

III．Lecturer

School of Silicon Valley Artisan of USJ Instructor: Yuting

(Master of Statistics from San Jose State University, proficient in Python, currently working in Gosvea, Inc Big Data Analysis group)

IV. Organizer

USJEdu

FREEDOM . TRUTH . SERVICE

Scan QR code to follow us

USJEdu Group