Kaggle Search competitions, datasets, kernels, discussion, and more... Sign In Register
Home Competitions Datasets Kernels Discussion Jobs Blog Forums Messaging Profile
Home Kernels, jeongbin and 2018 Kernels Discussions Datasets Competitions Jobs
jeongbin0 Jeong Bin Kim Follow 0 Votes 0 Forks 0 Comments
Notebook Data Logs Comments (0) Fork Notebook Version History Comments
Notebook
Data Sources
Kernel Output
Kernel Log
Comments (0)
Version 1 of 1 Created: (a few seconds ago) Runtime: 1 min 0 sec
Version History
Version 1 Created: (a few seconds ago) Runtime: 1 min 0 sec
Kernel Output
Output Files
Output Size: 0 bytes
Kernel Log
Log Size: 0 bytes
Comments (0)
No comments yet.
Kernel Info
Kernel Type: Notebook
Language: Python
Environment: Kernel
Kernel Id: 3

Title

2020년 데이터 분석을 위한 파이썬 기초 (BIG DATA 기초를 위한 파이썬 데이터 분석 입문)

Member

- 김진빈 (Jeongbin Kim)
- 김지원 (Jiwon Kim)
- 김지현 (Jihyeon Kim)

1. Objective

- 데이터 분석을 위한 파이썬 프로그래밍 입문 (Python for Data Analysis) 강의를 위한 기초 프로그래밍 예제

2. Dataset

- 2018년 교육데이터 분석 경진대회 - Jeonbuk Kaggle
- 데이터 출처 : 전북데이터허브
- 데이터 이용 조건 : 제한없음
- 데이터 제공 기관 : 전북특별자치도
- 데이터 제공 방법 : 파일 다운로드
- 데이터 갱신 주기 : 수시
- 데이터 최초 개방일자 : 2018
- 데이터 최종 개방일자 : 2018

3. Data Preprocessing

- 2018년도 교육데이터 분석 경진대회 - Jeonbuk Kaggle
- 데이터 출처 : 전북데이터허브
- 데이터 이용 조건 : 제한없음

1. 전북 교육데이터 분석 경진대회 (2018) 데이터 : 2015년 도내 중등학교 졸업자의 진로 현황 (Jeonbuk Kaggle 데이터 제공)
2. 전북 지역 고등학교 졸업자의 진로 현황 (전북데이터허브 제공)

- 전북 지역 고등학교 졸업자의 진로 현황 (전북데이터허브 제공) 데이터 분석 및 전처리 (Jeonbuk Kaggle 데이터 이용 : 2015년도 데이터 이용)

Unnamed:	시도명	시군명	학교명	학교구분	계열명	졸업자	진학자	취업자	기타
0	전북특별자치도	전주시	전주고등학교	일반고	인문계	320	300	15	5
1	전북특별자치도	전주시	전주여자고등학교	일반고	인문계	280	260	15	5
2	전북특별자치도	전주시	전주공업고등학교	특성화고	전문계	300	180	110	10
3	전북특별자치도	전주시	전주상업고등학교	특성화고	전문계	250	150	90	10
4	전북특별자치도	전주시	전주기전여자고등학교	특성화고	전문계	200	120	70	10

```
df_school_graduate.info()
```

```
<class'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 9):
Unnamed:    5 non-null int64
시도명         5 non-null object
시군명         5 non-null object
학교명         5 non-null object
학교구분        5 non-null object
계열명         5 non-null object
졸업자         5 non-null int64
진학자         5 non-null int64
취업자         5 non-null int64
기타          5 non-null int64
dtype: int64
memory usage: 1.1 KB
```

```
df_school_graduate.drop(['Unnamed:'], axis=1, inplace=True)
df_school_graduate.rename(columns={'시도명':'Sido', '시군명':'Sigun', '학교명':'School', '학교구분':'Type', '계열명':'Course', '졸업자':'Graduate', '진학자':'Advance', '취업자':'Employ', '기타':'Etc'}, inplace=True)
df_school_graduate.head(5)
```

Sido	Sigun	School	Type	Course	Graduate	Advance	Employ	Etc
전북특별자치도	전주시	전주고등학교	일반고	인문계	320	300	15	5
전북특별자치도	전주시	전주여자고등학교	일반고	인문계	280	260	15	5
전북특별자치도	전주시	전주공업고등학교	특성화고	전문계	300	180	110	10
전북특별자치도	전주시	전주상업고등학교	특성화고	전문계	250	150	90	10
전북특별자치도	전주시	전주기전여자고등학교	특성화고	전문계	200	120	70	10

```
df_school_graduate.groupby('Sigun').sum()
```

Sigun	Graduate	Advance	Employ	Etc
전주시	1350	1020	300	30

```
df_school_graduate.groupby('Type').sum()
```

Type	Graduate	Advance	Employ	Etc
일반고	600	560	30	10
특성화고	750	460	270	20

```
df_school_graduate.groupby('Course').sum()
```

Course	Graduate	Advance	Employ	Etc
인문계	600	560	30	10
전문계	750	460	270	20

```
df_school_graduate.groupby('Sido').sum()
```

Sido	Graduate	Advance	Employ	Etc
전북특별자치도	1350	1020	300	30

```
df_school_graduate.groupby('School').sum()
```

School	Graduate	Advance	Employ	Etc
전주고등학교	320	300	15	5
전주공업고등학교	300	180	110	10
전주상업고등학교	250	150	90	10
전주기전여자고등학교	200	120	70	10
전주여자고등학교	280	260	15	5

```
df_school_graduate.groupby('Type').sum()
```

Type	Graduate	Advance	Employ	Etc
일반고	600	560	30	10
특성화고	750	460	270	20

```
df_school_graduate.groupby('Course').sum()
```

Course	Graduate	Advance	Employ	Etc
인문계	600	560	30	10
전문계	750	460	270	20

```
df_school_graduate.groupby('Sido').sum()
```

Sido	Graduate	Advance	Employ	Etc
전북특별자치도	1350	1020	300	30

```
df_school_graduate.groupby('School').sum()
```

School	Graduate	Advance	Employ	Etc
전주고등학교	320	300	15	5
전주공업고등학교	300	180	110	10
전주상업고등학교	250	150	90	10
전주기전여자고등학교	200	120	70	10
전주여자고등학교	280	260	15	5

4. Data Visualization

- 2018년도 교육데이터 분석 경진대회 - Jeonbuk Kaggle
- 데이터 출처 : 전북데이터허브
- 데이터 이용 조건 : 제한없음

- 시군별 졸업자 수

```
plt.rcParams['axes.unicode_minus'] = False
ax = df_school_graduate.groupby('Sigun').sum()['Graduate'].plot.bar(rot=0)
plt.ylabel('Graduate')
plt.title('GRADUATE_SIGUN')
```

- 시군별 졸업자 수 대비 진학자, 취업자, 기타 현황

```
df_school_graduate.groupby('Sigun').sum()[['Graduate', 'Advance', 'Employ', 'Etc']].plot.bar(rot=0)
plt.ylabel('Graduate')
plt.title('Sigun')
```

- 학교별 졸업자 수

```
plt.rcParams['axes.unicode_minus'] = False
ax = df_school_graduate.groupby('School').sum()['Graduate'].plot.bar(rot=0)
plt.ylabel('Graduate')
plt.title('SCHOOL_GRADUATE')
```

- 학교별 졸업자 수 대비 진학자, 취업자, 기타 현황

```
df_school_graduate.groupby('School').sum()[['Graduate', 'Advance', 'Employ', 'Etc']].plot.bar(rot=0)
plt.ylabel('Graduate')
plt.title('School')
```

- 진학자 현황 (line)

```
plt.rcParams['axes.unicode_minus'] = False
ax = df_school_graduate.groupby('Sigun').sum()['Advance'].plot.line(rot=0)
plt.ylabel('Advance')
plt.title('ADVANCE_SIGUN')
```

- 취업자 현황 (bar)

```
plt.rcParams['axes.unicode_minus'] = False
ax = df_school_graduate.groupby('Sigun').sum()['Employ'].plot.bar(rot=0)
plt.ylabel('Employ')
plt.title('EMPLOY_SIGUN')
```

- 기타 현황 (line)

```
plt.rcParams['axes.unicode_minus'] = False
ax = df_school_graduate.groupby('Sigun').sum()['Etc'].plot.line(rot=0)
plt.ylabel('Etc')
plt.title('ETC_SIGUN')
```

5. Analysis based on Theory

- 이론적 분석 (Theory)

Strain Theory

- Strain Theory의 기원은 Durkheim의 아노미 이론에 기원하고 있다. 개인이 사회에 적응하는 방식에 대한 이론이다. 개인이 사회에 적응하는 방식에 대한 이론이다. 개인이 사회에 적응하는 방식에 대한 이론이