Introduction to Big Data

- Developed by Dr. Keungoui KIM
- https://awekim.github.io/portfolio/

Lecture 9. Inferential Analysis

```
from google.colab import drive
drive.mount('/content/drive')
```

```
import pandas as pd
import seaborn as sns
```

Simpson's Paradox

+ 코드 + 텍스트

```
data = pd.read_csv("/content/drive/MyDrive/[Lecture]/빅데이터개론/BigData_Python/09_InferentialAnalysis/uc
```

```
data.head()
```

```
data_sum = data.groupby(['Gender','Admit']).sum().reset_index()
data_sum
```

```
#data_sum_wide = data_sum.pivot_table(index=['Gender'],columns='Admit',values='n').reset_index()
data_sum_wide = data_sum.pivot(index=['Gender'],columns='Admit',values='n').reset_index()
#data_sum_wide = data_sum_wide[['Gender','Admitted','Rejected']]
data_sum_wide
```

```
data_sum_wide['total']=data_sum_wide['Admitted']+data_sum_wide['Rejected']
data_sum_wide
```

```
data_sum_wide['Admitted_prop'] = data_sum_wide['Admitted'] / data_sum_wide['total']
data_sum_wide
```

```
sns.barplot(x='Gender',y='Admitted_prop', data=data_sum_wide)
```

```
data.head()
```

```
#data_dept_wide = data.pivot(index=['Gender','Dept'],columns='Admit',values='n').reset_index()
data_dept_wide = data.pivot_table(index=['Gender','Dept'],
                                  columns='Admit',values='n').reset_index()
data_dept_wide
```

```
data_dept_wide['total']=data_dept_wide['Admitted']+data_dept_wide['Rejected']
data_dept_wide
```

```
data_dept_wide['Admitted_prop'] = data_dept_wide['Admitted'] / data_dept_wide['total']
data_dept_wide
```

```
sns.barplot(x='Dept',y='Admitted_prop', hue='Gender', data=data_dept_wide)
```

```
!pip install yfinance

import yfinance as yf

goog_df = yf.download('GOOG',start='2021-01-01',end='2021-12-31',progress=False)
goog_df.head()

goog_df.corr()
```