

Data Engineering

# 데이터엔지니어링

[ 실 세계 데이터셋에 대한 데이터 분석 ]



# 실 세계 데이터셋 분석



## 학습내용

### 1 실 세계 데이터셋 분석

## 학습목표

- 실 세계 데이터셋에 대한 간단한 데이터 분석을 수행할 수 있다.



# 데이터 분석 수행

총 줄 수 구하기

이벤트 수 구하기

최소 사람 ID 구하기

최대 사람 ID 구하기

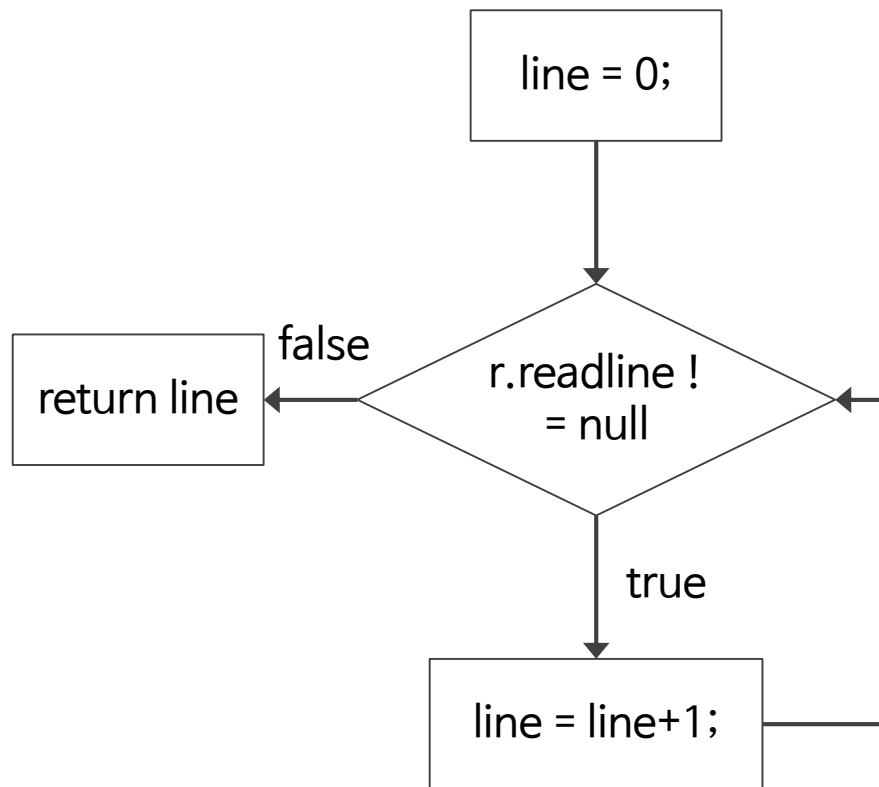
# 총 줄 수 구하기

d:\Wemail.txt 예제

```
# Directed graph (each unordered...  
# Email network of a large Europe...  
# Nodes: 265214 Edges: 420045  
# FromNodeId ToNodeId
```

0 1 0이 1에게 메일을 보낸 이벤트

0	2
0	4
1	3
2	5
3	1
3	4



데이터 총 줄 수 = 11



## 데이터 분석 수행

총 줄 수 구하기

이벤트 수 구하기

최소 사람 ID 구하기

최대 사람 ID 구하기

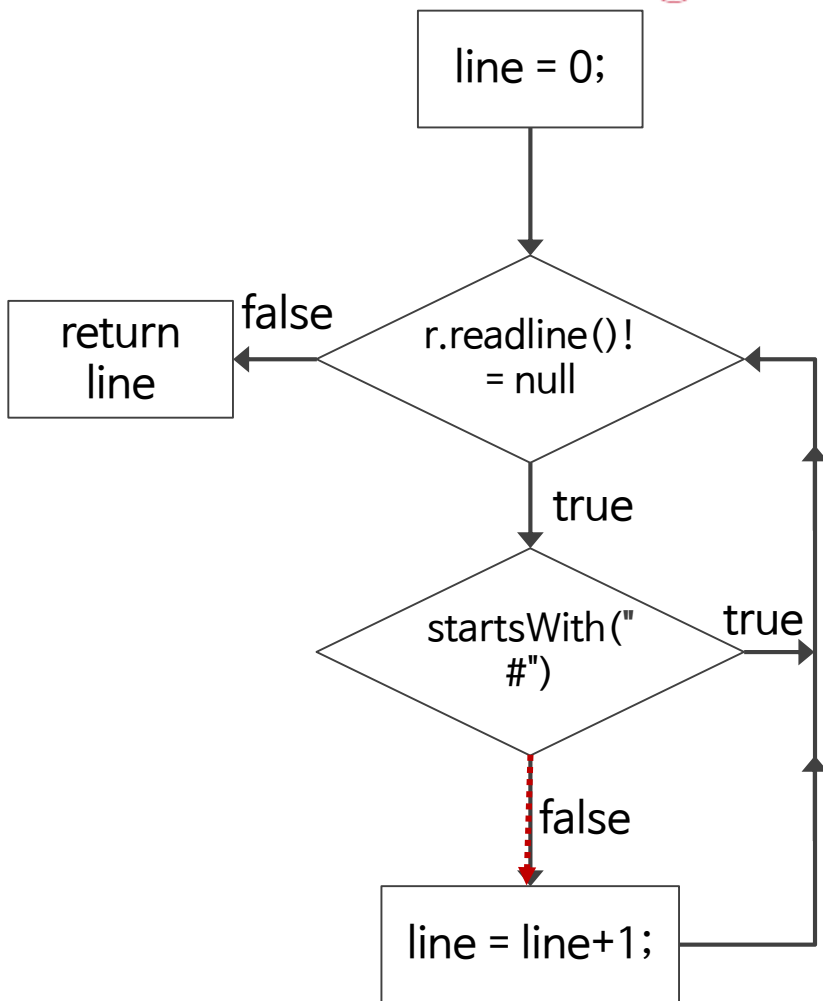
# 이벤트 수 구하기

d:\Wemail.txt 예제

```
# Directed graph (each unordered...  
# Email network of a large Europe...  
# Nodes: 265214 Edges: 420045  
# FromNodeId ToNodeId
```

0 1 0이 1에게 메일을 보낸 이벤트

0	2
0	4
1	3
2	5
3	1
3	4



이벤트 수 = 7



## 데이터 분석 수행

총 줄 수 구하기

이벤트 수 구하기

최소 사람 ID 구하기

최대 사람 ID 구하기



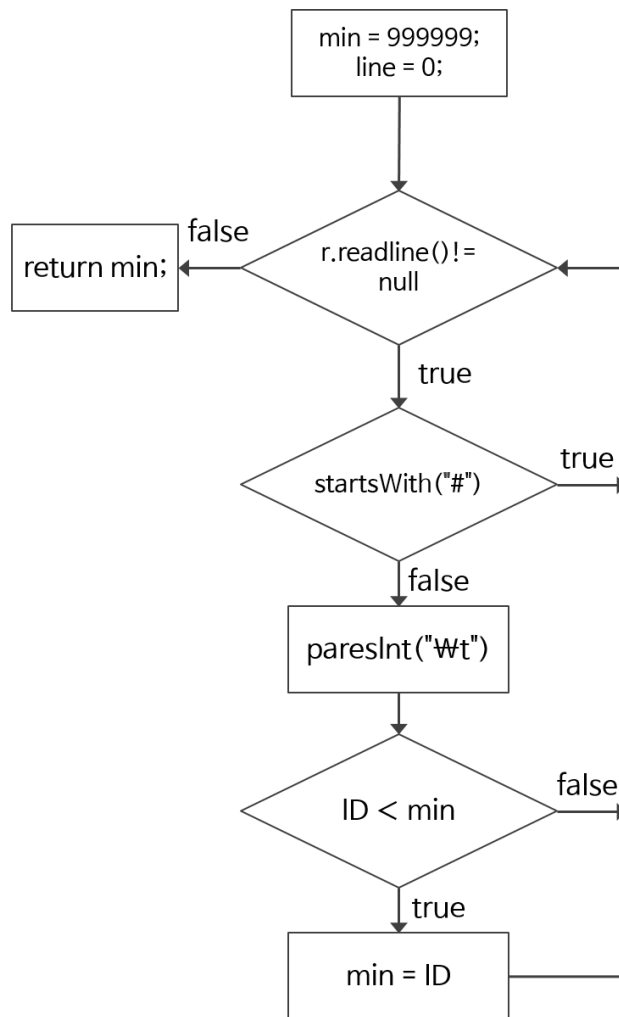
# 최소 사람 ID 구하기

d:\Wemail.txt 예제

# Directed graph (each unordered...  
# Email network of a large Europe...  
# Nodes: 265214 Edges: 420045  
# FromNodeId ToNodeId

0 1 0이 1에게 메일을 보낸 이벤트

0	2
0	4
1	3
2	5
3	1
3	4



모든 ID를 확인한 이후, Min은 **최솟값** 유지



## 데이터 분석 수행

총 줄 수 구하기

이벤트 수 구하기

최소 사람 ID 구하기

최대 사람 ID 구하기

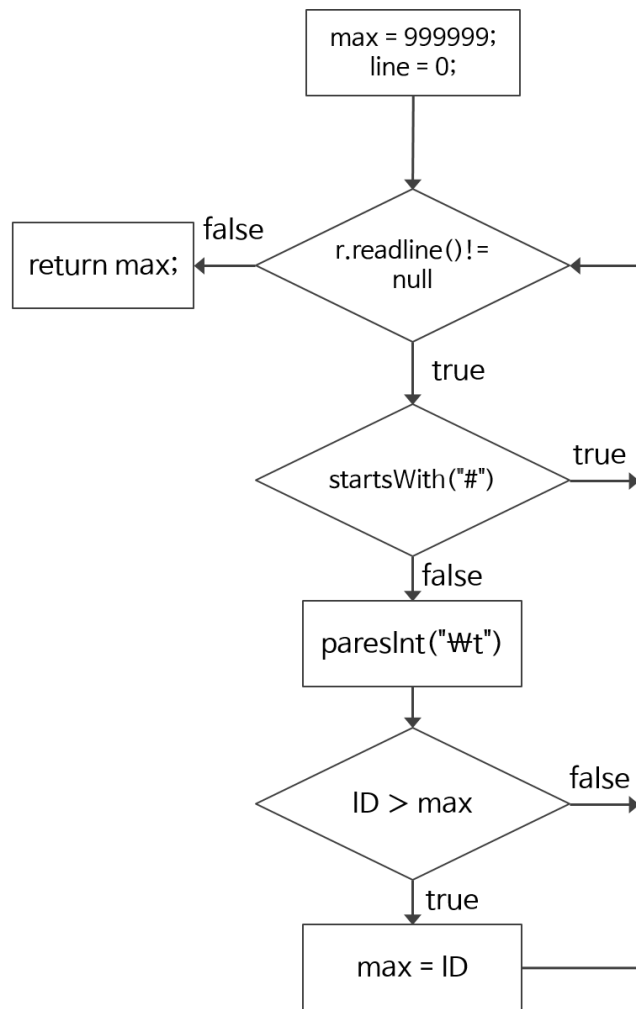
# 최대 사람 ID 구하기

d:\Wemail.txt 예제

# Directed graph (each unordered...  
# Email network of a large Europe...  
# Nodes: 265214 Edges: 420045  
# FromNodeId ToNodeId

0 1 0이 1에게 메일을 보낸 이벤트

0	2
0	4
1	3
2	5
3	1
3	4



# Remind

---

실 세계 데이터셋 처리 워밍업