

데이터 마이닝 프로젝트 제안서

3조

2014147576 박건우

2014147565 WANG GENGYU

2014147519 손장현

목차



문제 정의



관련 연구



방법론



일정

목차



문제 정의



관련 연구



방법론



일정

1

문제 정의

기상 조건에 따른
음식물 쓰레기 배출량 예측



1

음식물 쓰레기 배출의 53%를 가정에서 배출
음식물 처리는 일률적으로 처리됨
(부평구 기준, 주 3회)

53%



처리 비용



2

음식물 쓰레기 배출량 증가 시
빠른 수거가 불가능, 악취 민원 발생 및 음식
물 쓰레기통이 가득차는 경우도 발생
음식물 쓰레기 배출량 감소 시
다 차지 않은 음식물 쓰레기통 처리를 위해
불필요한 처리 비용 발생

4

음식물 쓰레기의 배출은 기상 조건에 따라 변동
가능하다 판단, 배출량 예측은 배출된 음식물 쓰레
기 수집 등 실생활에서 많은 활용 가능성이 있음

예측 필요



기상과 관련



3

제철 과일, 제철 요리 등 기후에 따른 식습관 변화
여름철 수박 등 과일 껍질 다량 발생
여름철 음식물 변질 심화
폭우, 폭설 등 악천후 시 배출 난해

1 문제 정의 – 기상 데이터

	Feature name		Feature name
Metadata	Region		Snow cover
	Date(Time stamp)		Snow cover(3h)
Observation	Temperature	Observation	Sky coverage
	Rainfall		Sky coverage(mid)
	Wind velocity		Cloud shape
	Wind direction		Cloud height(mx-mn)
	Humidity		Visibility
	Atmospheric pressure		Ground condition
	Dew point		Phenomenon index
	Spot atmospheric pressure		Ground temperature
	Sea-level pressure		Under(5cm) temperature
	Sunshine amount		Under(10cm) temperature
	Insolation		Under(20cm) temperature
			Under(30cm) temperature

지자체별 일 단위로 획득가능하여 지역적인 기상 조건 정보를 알 수 있다.

1 문제 정의 – 음식물 배출 데이터

항목명(영문)	항목명(국문)	항목크기	항목구분	샘플데이터	항목설명
count	총 갯수	INTEGER	1	10	목록 총 갯수
page	페이지번호	INTEGER	0	1	
rowNum	한 페이지당 row 갯수	INTEGER	0	10	
list	목록	INTEGER	0..n	-	정보 목록
disYear	배출년	CHAR(4)	1	2015	형식: YYYY
disMonth	배출월	CHAR(2)	1	01	형식: MM
disDate	배출일	CHAR(2)	1	01	형식: DD
disDay	배출요일	NUMBER(1,0)	1	4	형식: 1-7(일-월)

cityCode	지자체 영역코드	CHAR(3)	1	W34	
citySidoName	지자체 시도명	VARCHAR2(50)	1	경기도	
citySggName	지자체 시군구명	VARCHAR2(50)	1	화성시	
disQuantity	배출량	NUMBER(10,0)	1	74965700	단위: kg
disQuantityRate	배출량비율	NUMBER(12,2)	1	3.46	단위: %
disCount	배출횟수	NUMBER(10,0)	1	35679	
disCountRate	배출횟수비율	NUMBER(12,2)	1	3.35	단위: %

음식물 배출 데이터는 공공데이터 포털 (data.go.kr)의 [음식물 쓰레기 배출정보 서비스] api를 활용하여 획득하였다. 지자체/아파트 배출내역(일별, 요일별) 데이터를 획득할 수 있으며 해당 데이터는 다음과 같이 구성 된다.

- 3월~10월 : 수거전날 20:00 ~ 24:00

- 11월~2월 : 수거전날 19:00 ~ 24:00

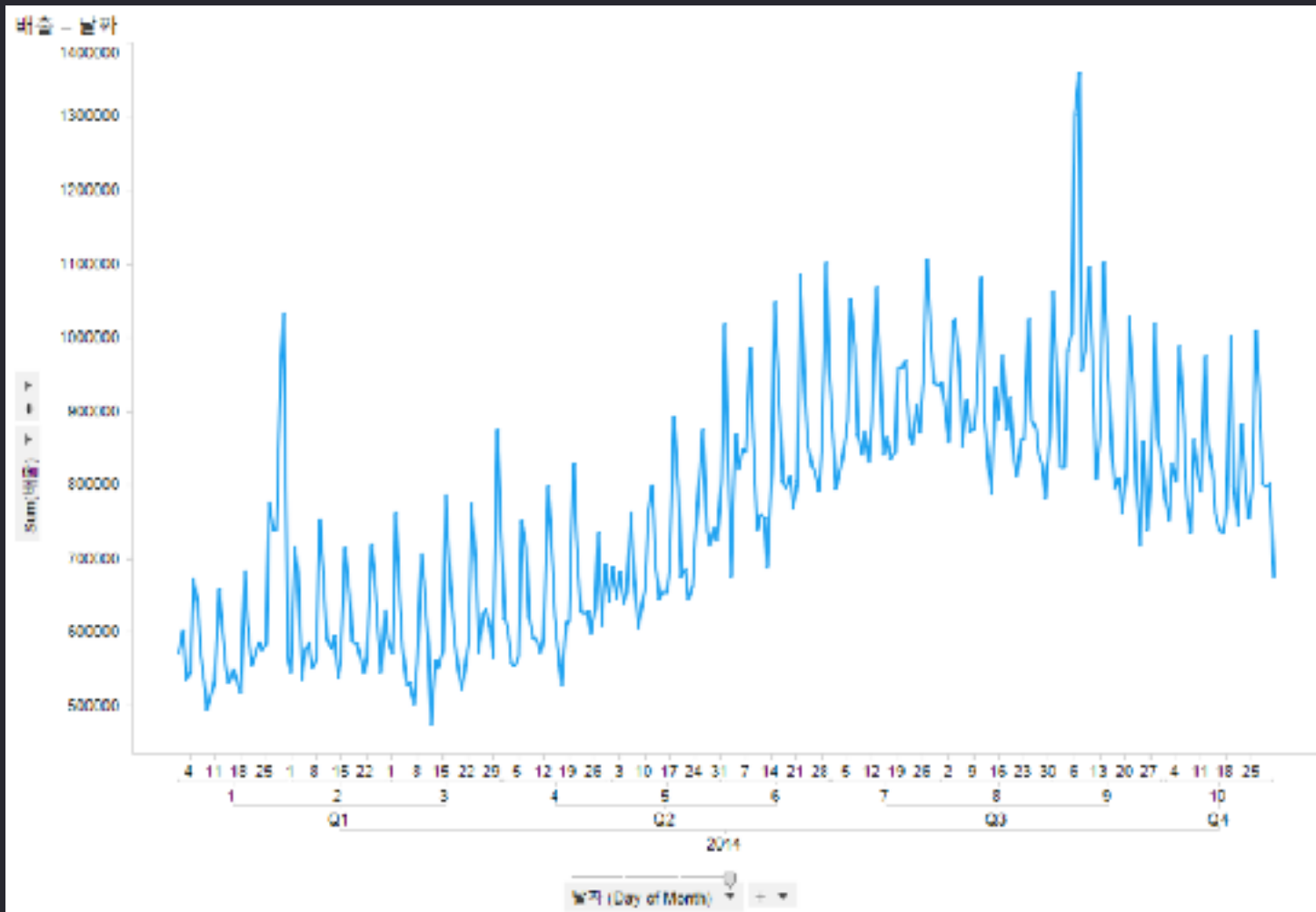
※단, 광복4차로 22:00~01:00

5. 이 봉투는 구로구 (신영동)에서만 사용하여야 합니다.

- 신도림동, 구로2동,

문제 정의 - 데이터 시각화

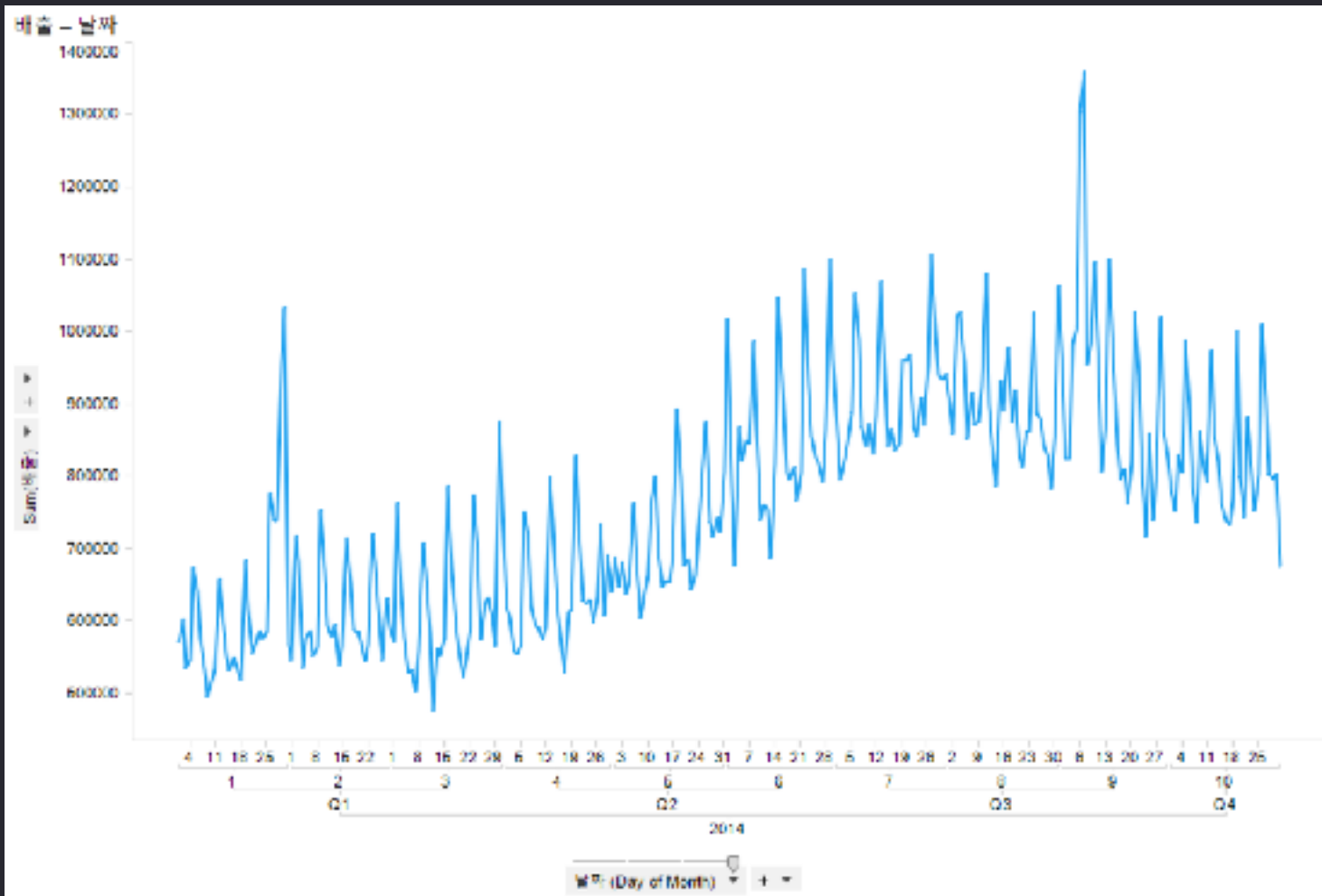
2014년도 일별 배출량



1월, 2월, 3월 등 겨울보다 7월, 8월, 9월의 여름의 배출량이 일반적으로 많은 것을 확인할 수 있다. 설과 추석 전후로 배출량이 급증하는 경우도 나타난다.

문제 정의 - 데이터 시각화

2014년도일별 배출 횟수

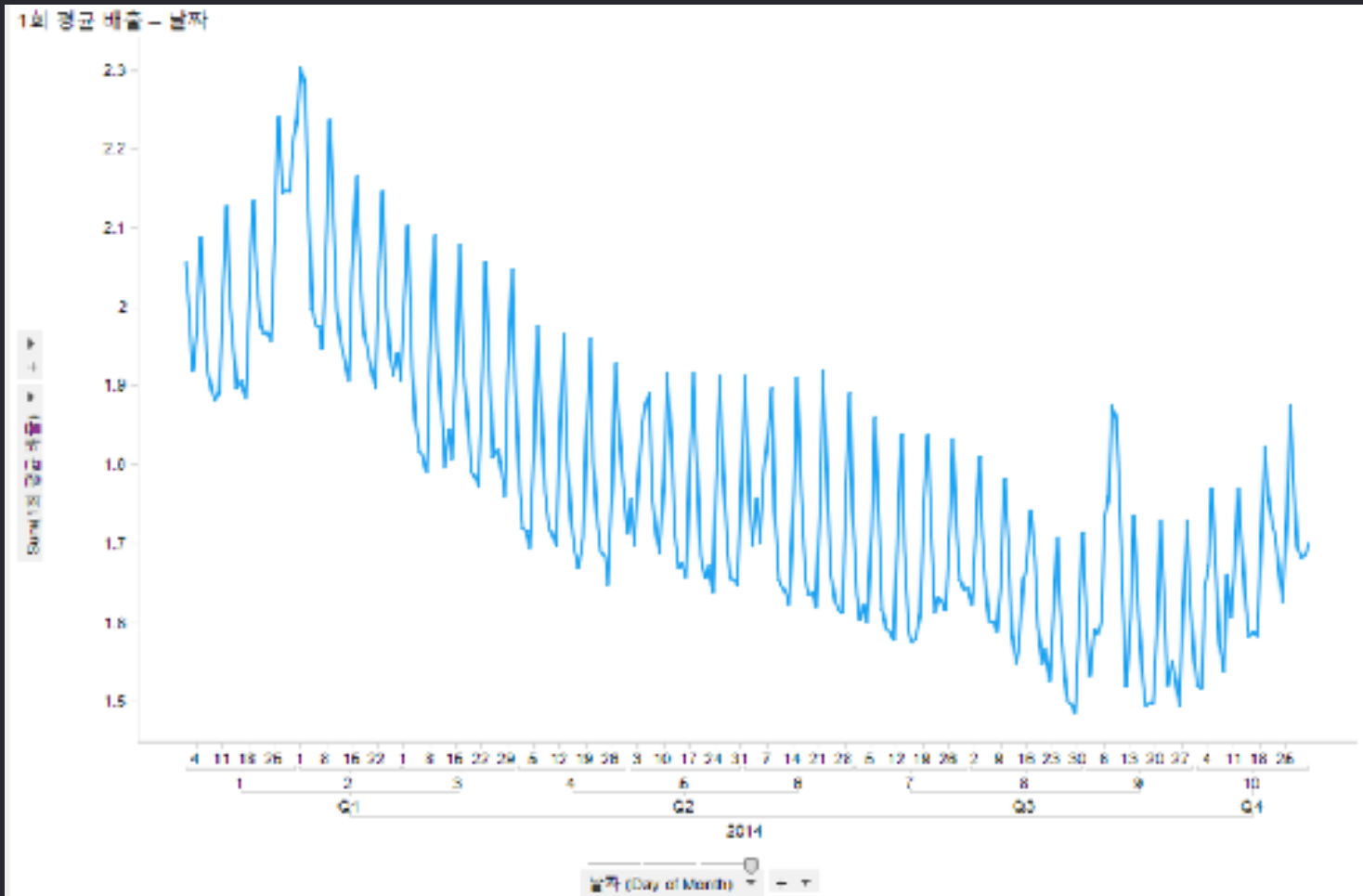


1월, 2월, 3월 등 겨울보다 7월, 8월, 9월의 여름의 배출 횟수가 일반적으로 많은 것을 확인할 수 있다.

설과 추석 전후로 배출 횟수가 급증하는 경우도 나타난다.

문제 정의 - 데이터 시각화

일별 배출량/배출회수의 값



한번 배출할 때 여름에 덜 배출한다.

여름철에 음식물이 더 빨리 상하여 가정에서 음식물 쓰레기를 보관하면 악취 등의 상황이 발생할 수 있기에, 음식물 쓰레기를 모아두지 않고 더 자주 배출한다고 추측해 볼 수 있다.

목차



문제 정의



관련 연구



방법론



일정

- 기상 조건 별 음식물 쓰레기 배출량 예측에 대한 기존 연구는 없다.
- 다른 요인을 활용한 음식물 쓰레기 배출량 예측에 대한 연구는 없다.
- 환경부에서는 이전의 배출량 데이터를 분석, 증가 정도를 파악하여 그것을 통해 앞으로 같은 정도로 증가할 것을 가정하고 예측하였다. 이 또한 일별 예측보다는 연도별 예측으로 진행한 것이다.

목차



문제 정의



관련 연구

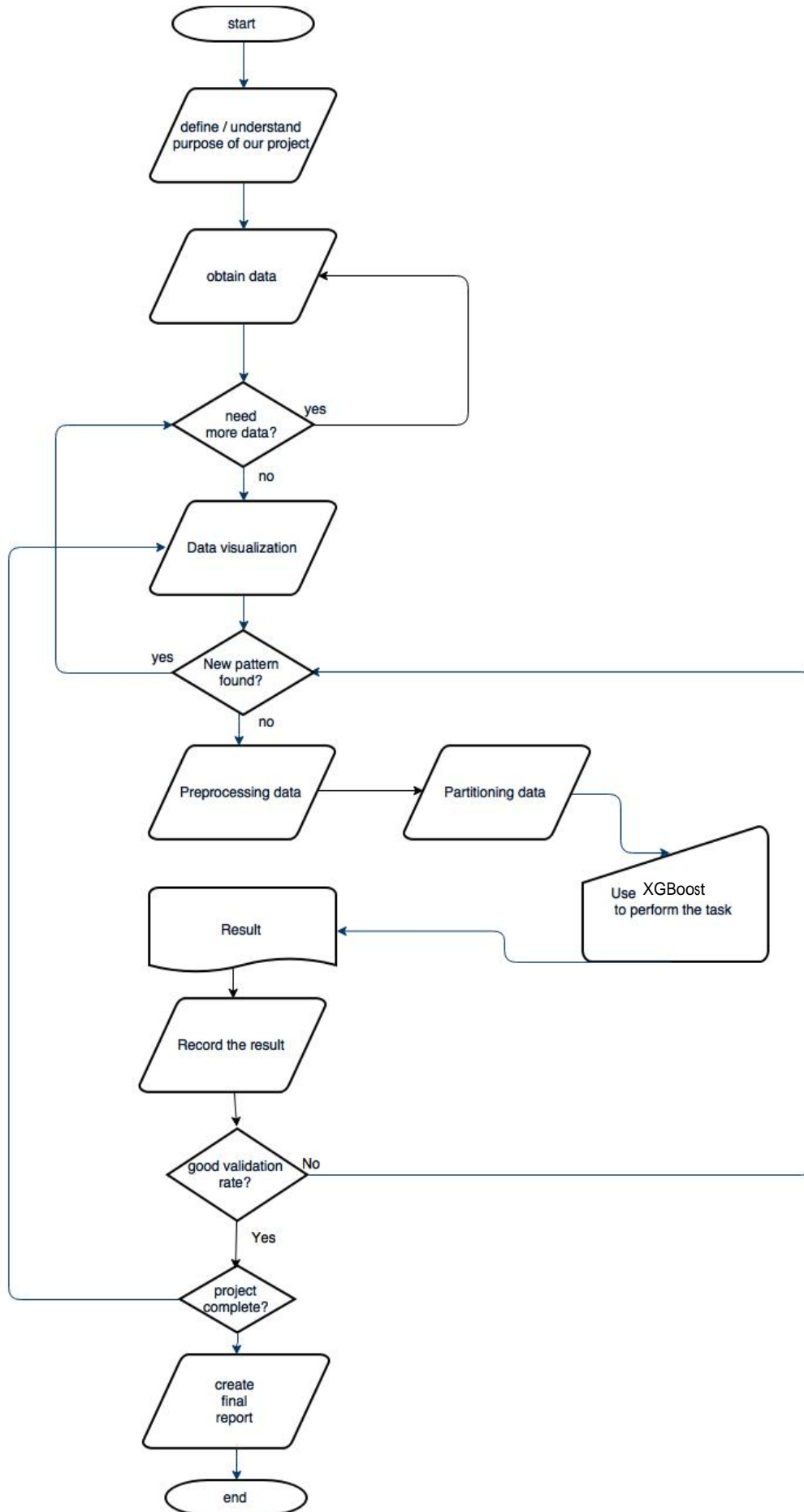


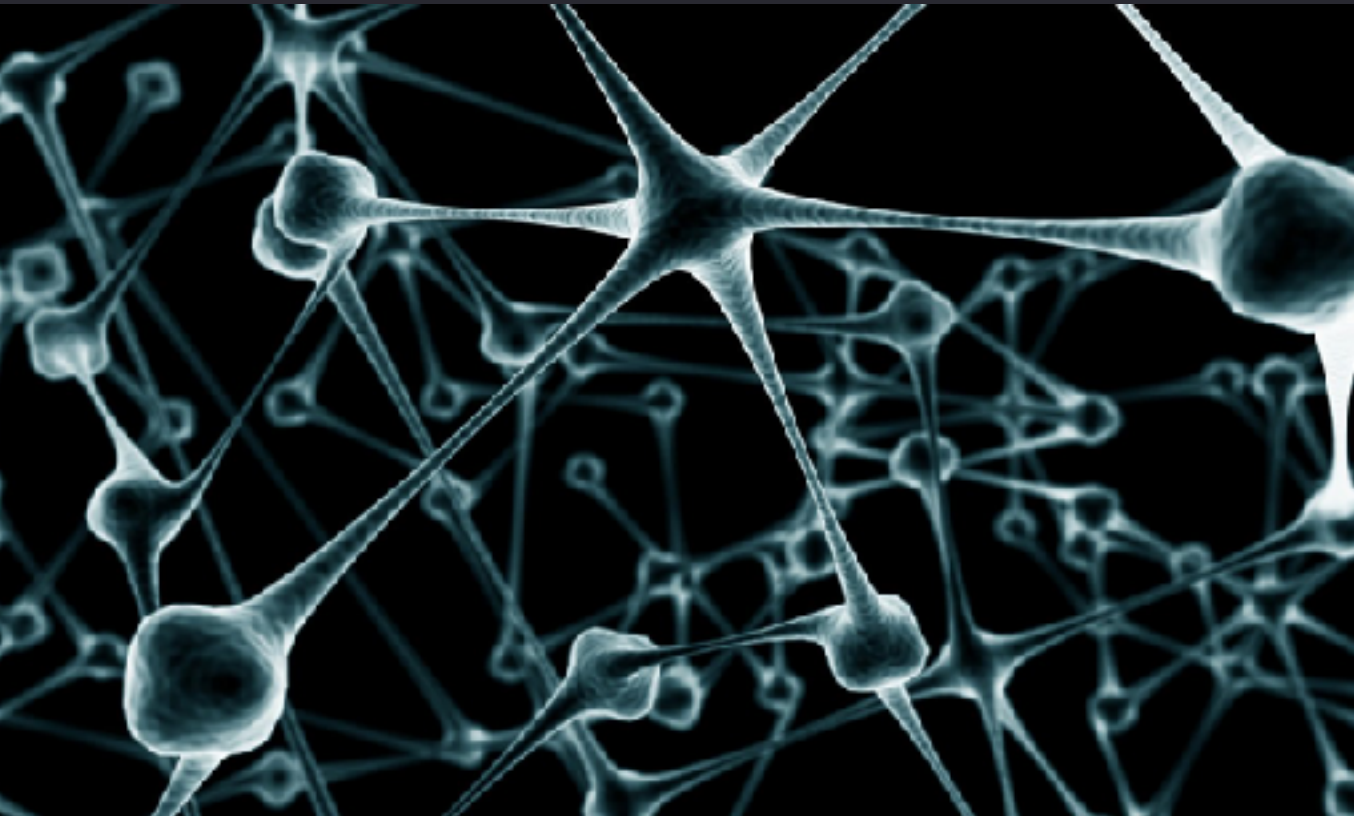
방법론



일정

<방법>





Xgboost 알고리즘의 input으로 사용하기 위해 data를 전처리하고 가공한다.



기온이나 강수량 정보는 normalization이 필요해보인다. 구름 많음, 맑음과 같은 날씨 정보는 dummy variable을 활용한다.



예측에 커다란 영향을 끼치는 feature와 그렇지 않은 feature를 제대로 구분하는 작업이 필요하다.

Preprocessing data

2

Method

90%
training
data

10%
test
data

Partitioning data

3

Method



Train Model

training by Xgboost



Test Model

test data 적용해서 validation
rate를 계산하다.



Result

좋은 결과값을 얻은 경우 상관관계를
분석해서 레포트에 작성

기대 효과

앞서 설명한 것처럼 관련 연구가 존재하지 않기 때문에 기상 조건과 음식물 쓰레기 배출량
사이에 주목할만한 좋은 상관관계가 존재하는 결과를 기대한다.

목차



문제 정의



관련 연구



방법론



일정

Project Schedule

일정 및 역할	4월 30일	~5월 4일	~5월 15일	~5월 27일	~5월 29일
왕경울	제안서 작성	주제 정리 추가 데이터 마련	Data mining 실행 평가 및 중간 보 고서 작성	반복적인 실행 평가에 따른 모델 tuning 결과 분석	결과 분석 결과 활용 방안 마련 보고서 작성
박건우	데이터 획득	데이터 mining 기 법 마련			
손장현	데이터 획득	데이터 pre- processing			

데이터 마이닝 프로젝트 제안서

Thank you