**2018년 2학기**

**다변량통계분석**

**기말 프로젝트**

**프로젝트 주제**

* 서울특별시 구별 쓰레기 관련 데이터 수집 및 분석을 통해

효율적인 쓰레기 처리 방안 제시

**프로젝트 주제 선정 배경**

1. 팀원들의 공통적인 경험으로, 이른 새벽 등교 혹은 늦은 시간 하교시, 쓰레기 수거차를 본

경험을 바탕

1. **쓰레기는 항상 배출이 되며, 배출된 쓰레기들은 지역별로 특징이 있지 않을까? 의문점**

**예) 각 지역별 인구거주 및 점포 특징 (명동주변지역 : 대량의 쓰레기 (관광객))**

1. 쓰레기 종류마다 특징이 있지 않을까? (ex : 일반쓰레기, 음식물쓰레기, 재활용쓰레기 등)
2. 쓰레기의 배출, 수집 및 처리 과정에 있어서 데이터분석을 통해 효율적인 쓰레기 처리 방안 탐색

**프로젝트 데이터 수집**

* 서울열린데이터광장 - ‘쓰레기’ 데이터 셋 (<http://data.seoul.go.kr/dataList/datasetList.do>)
* 서울열린데이터광장 – ‘지역구’ 데이터 셋 (<http://data.seoul.go.kr/dataList/datasetList.do>)
* 그 외 탐색 예정

**프로젝트 진행 시 예상되는 진행 방향**

1. 데이터 탐색

* 분석에 앞서 데이터를 파악하기 위한 탐색
* Pairs.Panels, Histogram, ggplot 등의 시각화 기법 사용

1. 데이터 전처리

* 데이터 분석을 진행하기 위한 데이터 전처리

1. 데이터 분석

* 회귀분석 : 변수간의 상관성 파악
* 군집분석 : 데이터의 군집 특성 파악
* PCA : 구별 특징을 파악

**프로젝트 구성 명단**

20132468 권오성, 20132605 안상준, 20132695 임필령, 20142560 이정훈