

DATA ANALYSIS PORTFOLIO

Introduction

Project1

Project2

Project3

Contact

About me

데이터 분석가, 황규상



안녕하세요. 데이터 분석가 황규상입니다.

2년차 데이터 분석가로 다양한 내/외부 데이터를 활용한 수요 예측 모델 개발 & 예측 파이프라인 개발 및 운영하는 경험을 통해 데이터 엔지니어링, 데이터 분석에 대한 역량을 키웠습니다.

Profile

경력 및 경험

한살림 2022.05 - 현재

- 분석 데이터 마트 구축
- 온라인 수요예측 모델 개발 및 운영
- 모델 성능평가 대시보드 개발 및 운영
- 차세대 ERP 시스템 개발 시 데이터 검증

Kdigital 융복합 프로젝트 과정 2021.06 - 2021.12

- 빅데이터 분석 교육 및 팀 프로젝트 진행
- Bigdata, AI, IoT, Cloud 팀 프로젝트
- 내부 경진대회 우수 프로젝트 수상

1인 광고 창작자 양성 과정 2021.06 - 2021.09

- 광고 기획, 촬영, 편집에 대한 기초 학습
- 팀 프로젝트 진행

학력

명지대학교 2014.02 - 2021.02

- 산업경영공학과 3.74 / 4.5

외국어

영어

- OPIC IM3 2022.03
- TOEIC 775 2022.03

보유 자격증

데이터 관련 자격증

- 빅데이터분석기사 2023.07
- SQLD(SQL 개발자) 2021.12
- ADsP(데이터분석전문가) 2020.07

보유 스킬

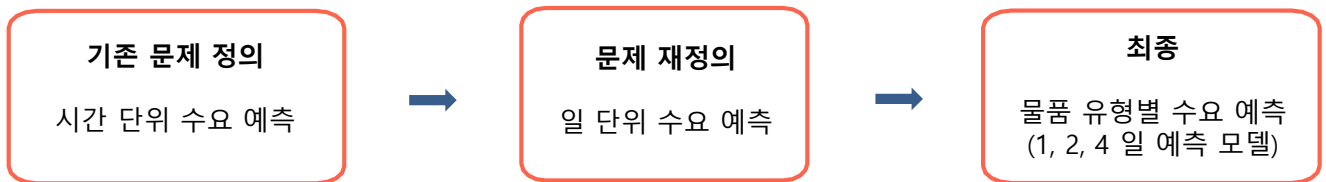
- Python
- SQL
- Tableau
- Google Analytics

Project 1

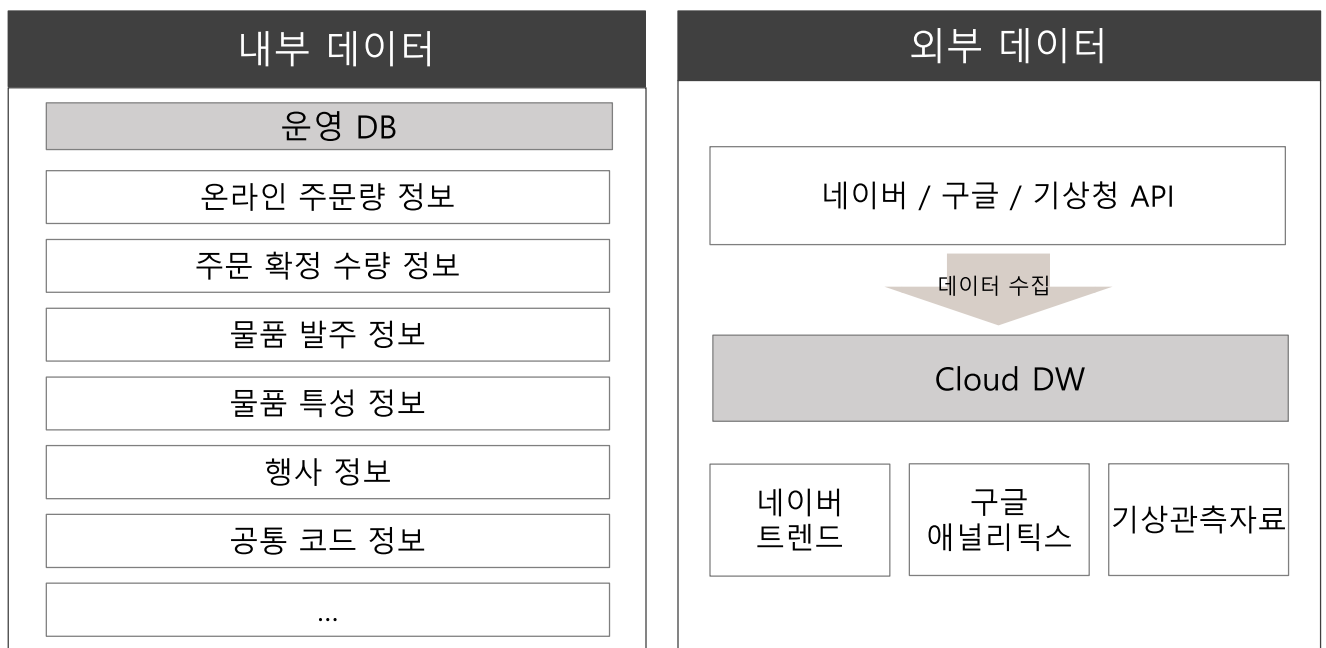
데이터분석 업무 중 주요 프로젝트

목적	온라인 수요예측 모델 개발 및 운영
기간	2022.06 - 2022.12
역할	문제 정의, 분석데이터 마트 구축, 데이터 추출/가공/모델링, 내부 도메인 지식 커뮤니케이션
Tool	SQL, Python
성과	<ul style="list-style-type: none"> 수요 예측 모델 개발 및 Daily 예측 파이프라인 개발 Daily 수요 예측을 위한 데이터 마트 및 인프라 구축 기존 수요 예측 정확도 약 20% 개선 온라인 리드타임 단축

문제 정의



활용 데이터



모델 개발

1. EDA, 데이터마트 구축

데이터 마트 구축

- EDA 통한 필요 데이터 정의
- 분석 데이터 마트 구축
- On-Premise DB <-> Cloud DB 연동 스케줄링

2. 데이터 전처리

전처리 필요 항목 정의

- 결측치, 이상치 처리
- 예측 기간 별 데이터 집계
- 비즈니스 로직 및 히스토리 반영(물품 정보 변경 등)
- 파생변수 생성(캘린더, 시계열 변수)

3. 모델링 대상 분류

적용 모델 분류

- 1차 모델링 후 정확도에 따른 품목 분류
- AutoML 모델링 품목
- 자체 Python 모델링 품목

4. 모델 앙상블

성격이 다른 두 모델 앙상블

- 주문량을 예측하는 주문량 모델
- 공급량을 예측하는 공급량 모델
- 두 모델 앙상블

모델 운영(수요 예측 파이프라인)

데이터 수집

데이터 가공

예측 결과 산출

예측 결과 집계

예측 결과
업로드

내/외부
데이터 수집
스크립트

데이터
전처리/
파생변수 생성
스크립트

모델 객체를 통한
예측 결과 산출
스크립트

예측 결과 집계
스크립트

DB 업로드
스크립트

Project 2

데이터분석 업무 중 주요 프로젝트

목적	모델 성능 모니터링을 위한 대시보드 개발
기간	2023.04 - 2023.06
역할	데이터마트 구축, 대시보드 개발 및 운영
Tool	SQL, Tableau
성과	<ul style="list-style-type: none"> • 모델 성능 관련 데이터 마트 구축 • 성능 모니터링을 위한 대시보드 개발 • 예측 전/후 대한 이슈 모니터링을 통한 즉각적인 대응 프로세스 수립

문제 정의

필요성

- Daily로 진행되는 예측의 성능 모니터링
- 현업에 성능 저하에 따른 이슈 사항 소통
- 모델 재학습 필요성 모니터링



대시보드 개발

- 예측 값 적재 전, 예상 성능 모니터링을 위한 대시보드 개발
- 예측 값 적재 이후, 현업에 이슈 사항 전달을 위한 대시보드 개발
- 모델 재학습 필요성 모니터링을 위한 대시보드 개발

활용 데이터

운영 데이터	분석 데이터마트
On-Premise DB	Cloud DB
실시간 주문 정보	예측 결과 정보
과거 주문 정보	네이버 트렌드
물품 발주 정보	구글 애널리틱스
물품 특성 정보	기상 관측 자료
행사 정보	
...	

대시보드 개발

1. 데이터마트 구축

데이터 마트 구축

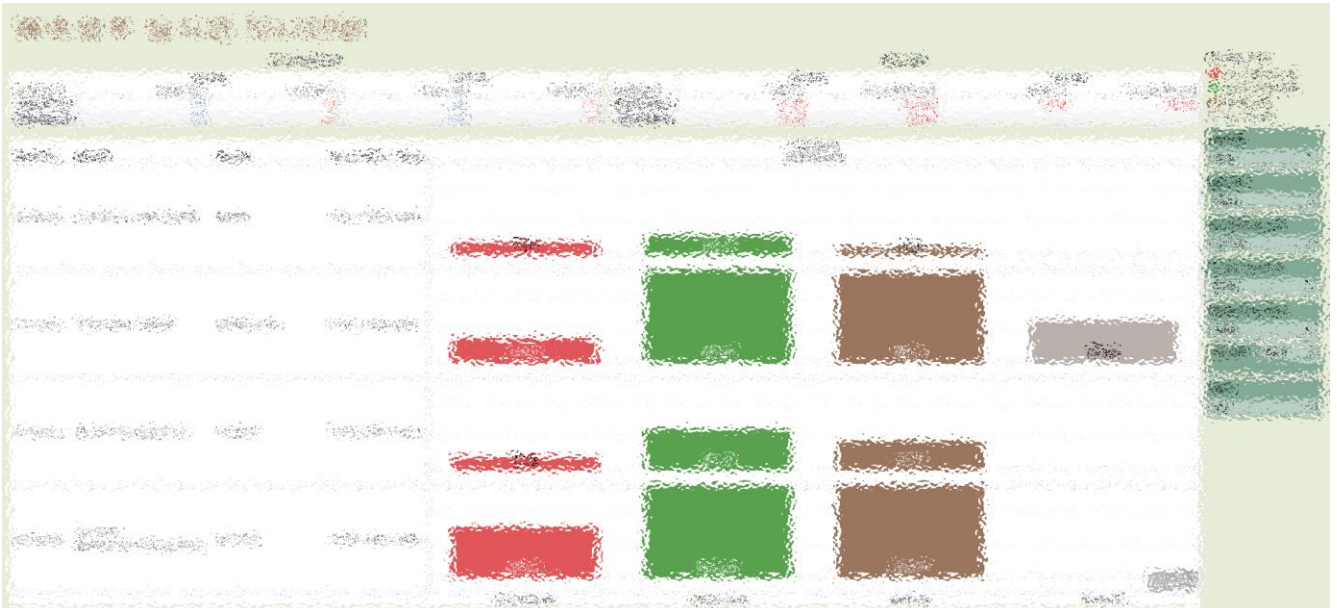
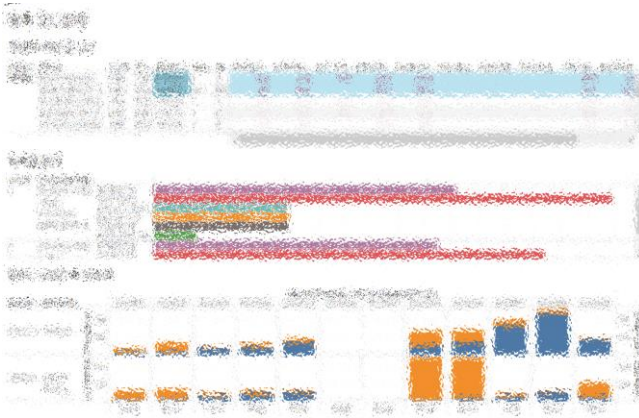
- 여러 테이블 활용하여 대시보드 개발에 필요한 데이터 마트 구축
- On-Premise DB <-> Cloud DB 연동

2. 대시보드 개발

유형 별 대시보드 개발

- 예측 값 적재 전 모니터링 대시보드
- 예측 값 적재 후 모니터링 대시보드
- 최근 예측 추세 모니터링 대시보드

대시보드 예시



Project 3

데이터 분석 공모전

목적	데이터 활용 아이디어&시각화 경진대회(서울특별시 배달 단속 지역 선정)
기간	2021.09 - 2021.11
역할	문제 정의, 데이터 추출/가공/모델링, 시각화, 통계 검증
Tool	Python
주최	과학기술정보통신기술부, 한국데이터산업진흥원, 데이터안심구역
성과	<ul style="list-style-type: none"> • 데이터 분석을 통한 서울특별시 배달 단속 지역 선정 • 우수 프로젝트 수상

문제 정의

사회적 이슈

- 코로나의 장기화로 인한 배달량 급증
- 배달량 급증으로 인한 이륜차 교통량 증가
- 이륜차 사고의 심각성
- 경찰청의 특별 단속 시행



아이디어

- 경찰청의 특별 단속 시행에 맞추어, 서울특별시 내의 단속 구역 선정
- 효율적인 단속에 기여

활용 데이터

내부 데이터

지역 생활 특성 데이터
편의시설 데이터
교통시설 데이터

지역별 가맹점 정보

지역별 매출 및 이용고객 정보

자택 직장 정보

외부 데이터

TAAS 교통사고 분석 시스템
공공데이터 포털

서/경기 이륜차 사고 데이터

전국 어린이 보호구역 표준 데이터

전국 노인&장애인 보호구역 표준 데이터

분석 과정

1. 가설 수립

가설의 시각적 검증

- 이륜차 사고 수에 지역별 특징이 있다는 가설 수립
- 서울의 지역별 이륜차 사고 수 지도 시각화
- 경기의 지역별 이륜차 사고 수 지도 시각화

2. 통계적 검증

가설의 통계적 검증

- 상관분석 : 지역별 사고 수와 다른 변수 상관관계
- 회귀분석 : 모델링을 통한 F통계량, R-squared 값, 변수별 p-value, 변수간 다중공선성 확인

3. 모델링

모델 개발 및 평가

- 랜덤포레스트 알고리즘을 활용 주요 변수 선정
- 경기 지역 데이터를 활용한 모델 개발
- 서울 지역 데이터를 통해 모델 평가

4. 문제 해결

법정동별 단속개수, 집중 단속 시간 추천

- 법정동 별로 주요 변수 활용, 최종 점수 도출
- 군집분석 통한 구 별 단속지역 개수 선정
- 법정동, 시간대 별 사고 데이터 시각화를 통해 집중 단속 시간대 추천

Contact & Channel

Email, Phone



Contact

성명 황규상, Kyusnag Hwang

생년월일 1995.09.12

연락처 010-5533-9783

Blog <https://kyusanghwang0912.github.io/>

Email rbtkd912@naver.com

주소 경기도 군포시 산본동