

- 목차 -

1. 프로젝트 개요

- 선정이유
- 프로젝트 목적

2. 프로젝트 진행절차

- 데이터 전처리
- 데이터 전처리 목적 및 필요성
- 데이터 전처리 방법 및 과정

3. 프로젝트 결과

- 도시
- 배달
- 결제

4. 프로젝트 제안

- 도시
- 배달
- 결제

5. 문제해결 과정 및 개선점

- 문제
- 해결방안
- 개선점

1. 프로젝트 개요

1) 선정이유

배달 플랫폼을 구성하는 다양한 이해관계자와 변수 간의 관계를

2) 프로젝트 목적

해외 딜리버리 앱 데이터 분석을 통하여 배달 플랫폼의 영업 총 이익에 영향을 주는 요소를 분석, 최대 효율을 창출할 수 있는 변수를 조사해본다

2. 프로젝트 진행 절차

1) 데이터 전처리

- 데이터셋 구조 : 주 데이터 1개와 보조데이터 6개 / 주데이터에 추적용 ID로 보조데이터와 연동됨
- 용어 정리

영문	국문	영문	국문
order_id	주문 고유 번호	hub_id	허브 고유 번호
order_status	주문 처리 상태	hub_name	허브 이름
order_amount	주문 금액	store_id	가게 고유 번호
order_delivery_fee	배달 금액	store_name	가게 이름
order_delivery_cost	배달 플랫폼 이용료	store_latitude/longitude	가게 위치 좌표값
order_moment_	주문 단계별 타임스탬프	hub_latitude/longitude	상권 위치 좌표값

2) 데이터 전처리 목적 및 필요성

- 문제점 : 각 팀이 데이터를 임의로 가공할 경우, 용어 및 데이터 일관성에 문제가 생김
- 목적 : 1) 소통의 원활함을 위해 데이터 통합 처리를 함
2) 파일용량 줄이기

- 성과 : Column 명, 수치 등 데이터 전반의 용어가 통일되며 이로 인해 소통 및 협업이 원활해짐

3) 데이터 전처리 방법 및 과정

- 1) 주 데이터와 각 보조데이터를 상호간의 Tracking ID를 이용하여 취합
 - 2) 과반의 동의를 얻어 사용하지 않는 컬럼을 삭제
 - 3) 총 4개 도시 중 관측데이터 수가 적은 2개 도시 제거
- 결측값 변경 :
 - payment method만 있는 행들은 : 배달이 아닌 pick up으로 간주
 - Payment method count에 활용할 수 있는 데이터로 삭제 보다는 driver_id컬럼에 '0'값을 입력하여 활용
 - 데이터(order_amount) 기준 분포확인 : 데이터 대부분이 왼쪽으로 치우쳐져 있으며 오른쪽으로 긴 꼬리 형태 보임.
 - payment amount도 같은 모양이라 큰 차이 없음
 - Box plot 으로 확인하여 이상치 제거
 - QQ plot 및 통계분석으로 정규분포에서 이탈하는 꼬리부분을 확인
 - Normality Tests for Order Amounts:
 - Shapiro-Wilk Test: statistic=0.924, p-value=0.000
-> p-value < 0.05 거절
 - Kolmogorov-Smirnov Test: statistic=0.106,
p-value=0.000
-> p-value < 0.05 거절
 - D'Agostino's K-squared Test: statistic=25428.236,
p-value=0.000 -> p-value < 0.05 거절
 - Skewness: 0.799 right-skewed / long right tail
 - excess kurtosis : -0.231 platykurtic에 가까움

3. 프로젝트 결과

1) 도시 :

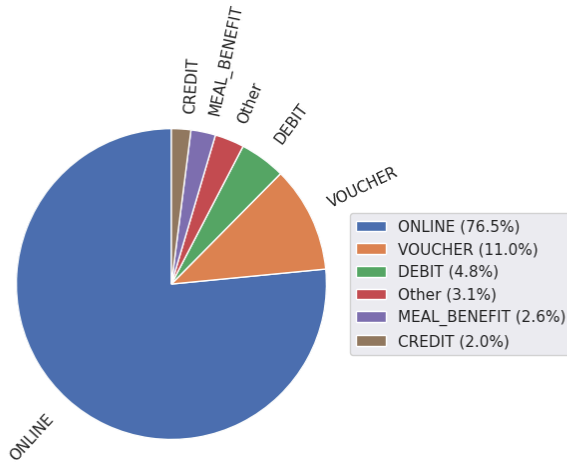
- hub-store 간 거리와 매출과는 관련이 없음
- 단, 각 hub 별로 종속 store들의 허브와의 거리가 대체로 일정함 : 일정한 곳에 끼리끼리 분포해 있음
- 비슷한 매출을 올리는 다수의 식당/매점과 높은 매출을 올리는 소수의 식당/매장으로 구성됨 /높은 매출을 올리는 식당도 항상 클러스터 안에 있음
- hub로 부터 멀리 떨어진 식당은 높은 매출을 올리지 않음

2) 배달:

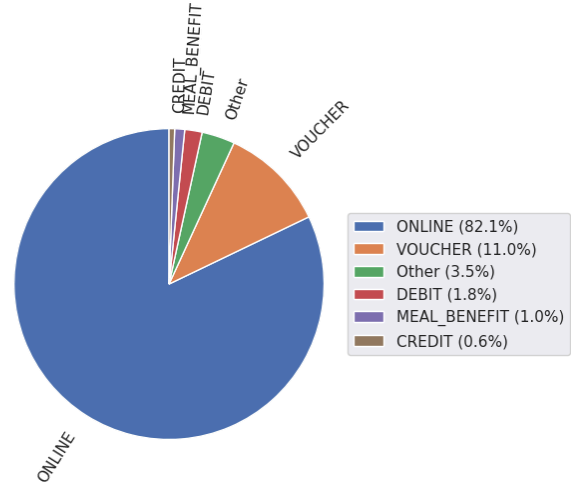
- 오토바이를 이용한 배달 건이 약 70%로 대다수의 배달기사가 자전거보다 오토바이를 이용하고 있음.
- 특히, 상근직 배달기사들은 자전거를 거의 이용하지 않음.
- 전체 배달 기사 중 상근직 비중이 20%가 채 되지 않으며 대다수의 배달건은 비상근직 기사들에 의해 완료됨.
- 배달 주문 금액과 이동 거리간 강한 양의 상관관계를 보임.
- 배달비는 배달 거리에 비례하여 측정됨.

3) 결제 방식:

Payment Method Distribution in RJ



Payment Method Distribution in SP



- 두 도시 모두 가장 많이 사용되는 결제 방식이 ONLINE이며, 두번째는 VOUCHER임
- Rio de Janeiro의 경우 상위 2개 항목이 87.5% 를차지, Sao Paulo의 경우 상위 2개 항목이 93.1%를 차지하며, 두 도시를 합하면 상위 2개 항목이 90.5%를 차지함

payment_method payment_fee			payment_fee	
0	VOUCHER	0.0	count	36448.0
14	VOUCHER	0.0	mean	0.0
25	VOUCHER	0.0	std	0.0
33	VOUCHER	0.0	min	0.0
37	VOUCHER	0.0	25%	0.0
...	50%	0.0
328848	VOUCHER	0.0	75%	0.0
328853	VOUCHER	0.0	max	0.0
328874	VOUCHER	0.0		
328875	VOUCHER	0.0		
328883	VOUCHER	0.0		

36448 rows × 2 columns

- VOUCHER로 결제한 경우, 결제 수수료가 0임을 확인

4. 프로젝트 제안

1) 도시

- 허브관리자 선정 : 각 허브마다 관리자를 따로 지정해서 관리하고, 매출이 높은 매장은 vip로 특별관리하여 promotion 혜택, payment fee
- 추가연계 사항 : 허브별 매출이 높은 가게의 주문금액 당 배달 거리를 산출하여, driving성과가 좋은 전담 motorbike driver를 배치

2) 배달

- 운송수단으로 오토바이가 압도적으로 높은 비율을 차지하고 있으나 전체 비중은 아님. 자전거 이용 중인 상근직 기사들에게 오토바이 도입 적극 권장하여 거리 대비 배달 시간 단축.
- 배달기사가 배달 완료 확인을 놓쳐 생긴 결측치가 적지 않음. 배달 확인 절차를 자동화하는 등의 방법을 도입하여 데이터 정확도.
- 원거리 주문건일수록 배달비 수수료 수익 역시 높아지므로 자전거 대신 오토바이를 도입하여 수익 증대를 기대할 수 있음.
- 오토바이 도입시 더 많은 원거리 배달건, 즉, 고가 배달건 소화가 가능해지므로 허브 분포 현황과 운송 수단 현황을 파악하여 배달 역량 강화

3) 결제 방식

- 결제 방법의 90% 가까이 차지하는 ONLINE 방법과 VOUCHER 방법에 중점을 두어 promotion 진행을 통한 앱 사용 유도
- ONLINE 결제 방식
 1. 일정 금액 이상 결제시 추가 할인
 2. 결제 금액 일부 포인트 적립
- VOUCHER 결제 방식
 1. 지역화폐 시스템처럼 앱에서 VOUCHER 충전시 일정 비율 추가 충전
 2. VOUCHER로 결제시 주문금액 일부 할인

5. 문제 해결 과정 및 개선점

1) 문제

- 데이터 양이 방대하고 지나치게 세분화되어있어 사용하기 좋은 데이터를 선별하고, 선별된 데이터를 다시 조합하는데 시간이 많이 소요됨
- 결측값에 대한 전처리가 많이 미흡하여 분석을 하는데 있어 정확한 결과가 나오지 않을 가능성이 존재

a) 해결방안

2) 개선점

- a. 처음 데이터 전처리시 꼬리부분(소수의 높은 매출)을 삭제하면 안되었음
- b.