

미국 카페 매출데이터 분석을 통한 국내 브랜드 해외 진출 방안

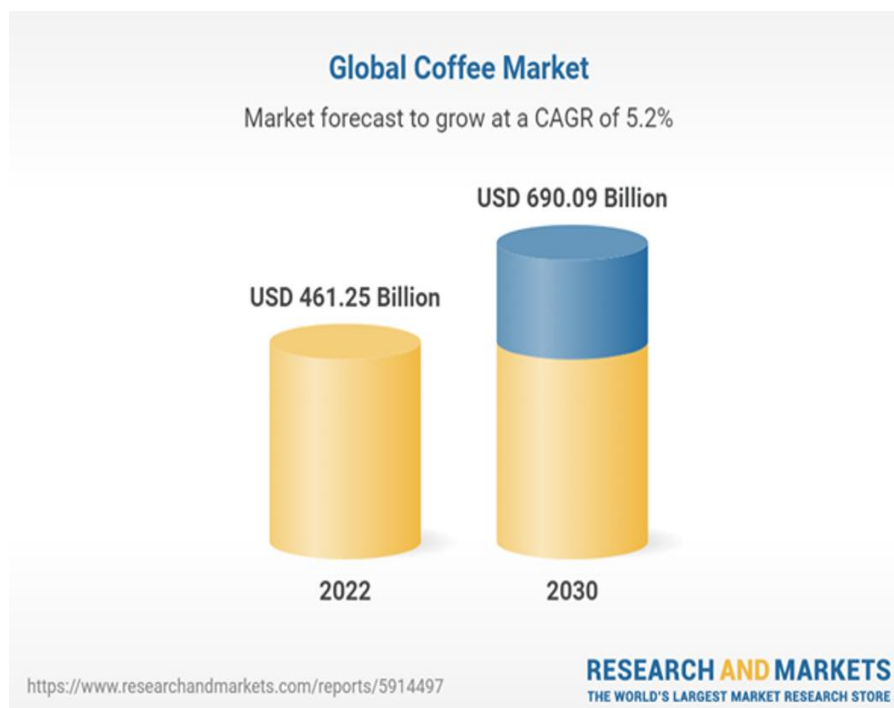
데사고팀 2조

황채원, 옥예진, 엄대식, 조수진

프로젝트 주제 선정 배경 및 필요성	3
데이터 소개.....	5
주요 분석 결과.....	7
지점 및 매출 목표 달성도 분석 (엄대식).....	7
주요 상품 데이터 분석(옥예진).....	16
재방문 데이터 분석(황채원).....	25
회원 데이터 분석(조수진)	34
프로젝트 결론.....	45
한계점	49

프로젝트 주제 선정 배경 및 필요성

최근 몇년간 글로벌 커피 산업은 전 세계적으로 급격한 성장을 이뤄냈으며, 특히 미국 시장은 그 중심에 있다. 미국은 세계 최대 커피 소비국 중 하나로 연간 약 1,460억 달러 규모의 커피시장을 형성하고 있고, 이는 전 세계 커피 소비량의 약 25%를 차지한다. 스타벅스와 같은 대형 체인부터 소규모 개인 카페까지 다양한 사업장들이 성공적으로 운영되고 있고, 이러한 시장 환경은 커피 산업이 여전히 큰 성장 잠재력을 지니고 있음을 시사한다.



[그림1] Global coffee market size forecasts 2023-2030 report

한편, 국내 커피 시장은 이미 성숙기에 접어들며 포화 상태에 이르렀다. 한국은 1인 당 연간 353잔의 커피를 소비할 정도로 높은 소비율을 보이고 있고, 국내 커피 시장은 이미 7조원 규모에 달한다. 이러한 상황 속 '해외로의 진출'은 새로운 성장 동력을 찾기 위한 과제로 떠오르고 있다. 특히, 미국과 같은 대규모 시장은 국내 브랜드들이 도전할 만한 가치가 있으며, 성공적인 진출을 통해 브랜드 가치를 글로벌 수준으로도 끌어올릴 수 있다.

그러나 해외 시장 진출은 기업 입장에서 리스크가 크다. 이에 실제 뉴욕에서 운영중인 커피 브랜드의 매출 및 고객 데이터 등을 활용하여 데이터 기반 전략적 접근을 통해 해당 지역의 시장 트렌드와 소비자 선호도 등 인사이트를 도출하고, 현지 브랜드와 차별점을 둘 수 있는 우리 기업이 취할 수 있는 전략을 제시하고자 한다.

본 보고서는 미국 카페 매출 데이터 분석을 통해 국내 커피 브랜드가 해외, 그 중 특히 미국을 대표하는 뉴욕 시장을 중심으로 성공적으로 진출할 수 있게 전략을 제안하고자 한다. 이는 국내 커피 산업의 글로벌 경쟁력을 강화하고, 새로운 성장 기회를 모색하는데 기여할 것이다.

데이터 소개

1. 사용 데이터셋 소개

Kaggle에 있는 'Coffee shop sample data' 데이터셋을 사용하였다.

링크: <https://www.kaggle.com/datasets/ylchang/coffee-shop-sample-data-1113/data>

해당 데이터셋은 2019년 4월에 누적된 카페의 매출 데이터, 상품 정보 데이터, 고객 데이터, 매장 정보 데이터 등을 포함하고 있다.

주요 분석 결과

지점 및 매출 목표 달성도 분석 (업대식)

1. outlet 지역 별 특성 조사 (도메인)

해당 데이터 셋은 뉴욕 내 세 지역의 매출 데이터 셋을 다룬다.

편의상 'sales_outlet_id' 의 3, 5, 8 로 지역을 호칭하여 지역별 특성을 조사해보았다.

sales_outlet_id	Store_address	sales_outlet_ty...	store_city	store_state_province	Neighborhood
3	32-20 Broadway	retail	Long Island City	NY	Astoria
5	100 Church Street	retail	New York	NY	Lower Manhattan
8	687 9th Avenue	retail	New York	NY	Hell's Kitchen

[도표1] SALES_OUTLET 별 위치

1.1 (3) Astoria

(3) 지역은 상업지역과 주거지역이 혼합된 지역으로, 다양한 문화적 배경을 가진 인구가 밀집해 있다. 유럽 여러 국가에서 온 이주민들이 모여 살고 있으며, 독일, 그리스, 이탈리아 등 수많은 문화권의 상점 및 레스토랑들이 밀집한 장소로 알려져 있다. 젊은 직장인들과 예술가들이 많이 밀집해있다.

1.2 (5) Lower Hanhattan

(5) 지역은 미국 최대의 증권거래소가 있는 '월스트리트', 명품 매장과 맛집 등이 즐비한 '소호' 등의 관광명소로 유명하고, 많은 대기업의 본사와 사무실이 밀집해 있다. 추가적으로, '자유의 여신상' 로 가는 페리를 탈 수 있는 장소로 많은 여행자들이 찾는다.

1.3 (8) Hell's Kitchen

(8) 지역은 센트럴파크나 브로드웨이 등 뉴욕의 주요 명소와 가까워 도시 관광을 위한 거점으로 유명하다. 주요 고객층은 관광객, 연극 관람객, 지역 주민으로 저녁시간대 및 주말에 활동이 많다.

세 지역의 특징을 요약하여 정리하면 다음과 같다.

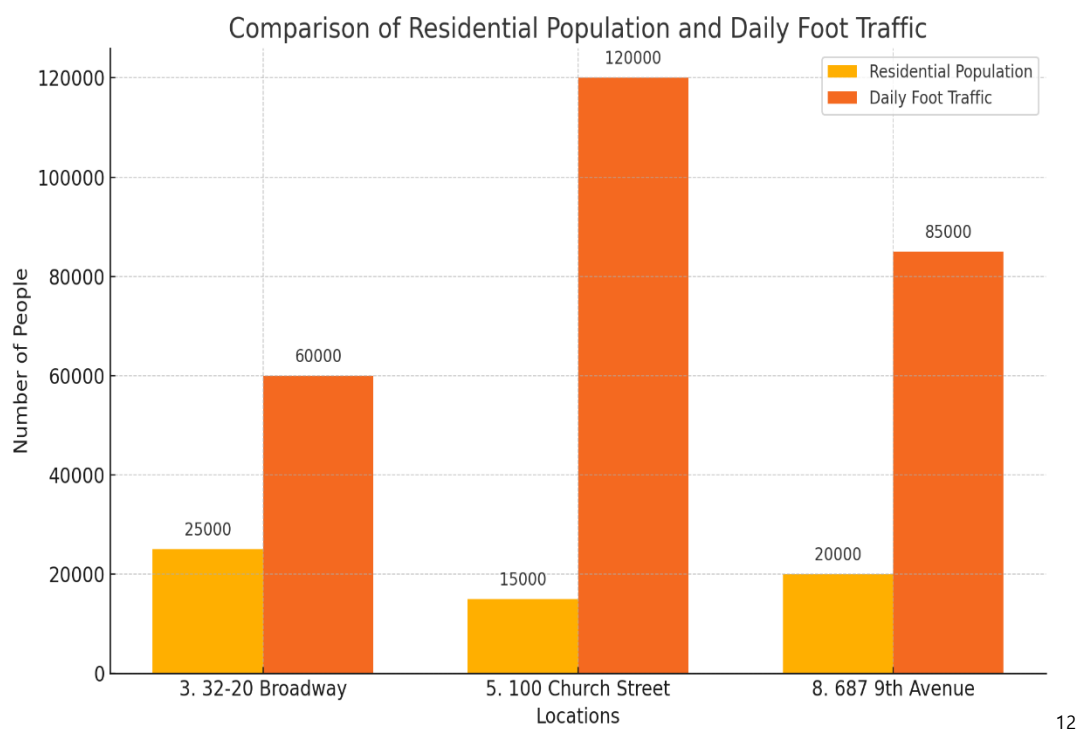
Outlet 3: 유럽 이주민 거주지역, 수많은 문화권과 상점

Outlet 5: 관광 명소 및 사무 지역

Outlet 8: 주요 명소 관광을 위한 거점

2. 거주인구 및 유동인구 분석 (도메인)

지역 별 특성을 알아본 후 각 지역 별 거주인구와 유동인구 추정 수치를 조사했다.



[차트1] 지역별 거주 인구 및 유동 인구 추정 현황

¹ 출처: NYC Open Data

<https://data.cityofnewyork.us/Transportation/Bi-Annual-Pedestrian-Counts/2de2-6x2h>

Outlet 3 지역의 거주인구는 25,000명, 유동인구 60,000명으로 거주인구는 비교적 많고, 유동인구는 적은수치로 보인다.

가장 높은 주거인구: 25,000명

가장 낮은 유동인구: 60,000명 가량 추정

Outlet 5 지역의 거주인구는 15,000명, 유동인구는 120,000명으로 비교군 중 거주인구는 가장 적으나, 유동인구는 가장 높은 수치를 보인다.

가장 낮은 주거인구: 15,000명

가장 높은 유동인구: 120,000명 가량 추정

Outlet 8 지역의 거주인구는 20,000명, 유동인구는 85,000명으로 유동인구와 거주인구 둘 다 중간 값에 해당한다.

중간 수준의 주거인구: 20,000명

중간 수준의 유동인구: 85,000명 가량 추정

3. 매장 별 판매 목표 및 카테고리 별 달성도 분석

매장 별 Goal 달성도 도출

단위: qty

sales_outlet_id	total_goal	total_goal_achieved	RATE	달성여부
3	18000	22897	127%	O
5	25000	22463	90%	X
8	22500	22932	102%	O

[도표2] 매장별 GOAL 달성도

‘Sales targets’ 테이블과 ‘Sales receipts’ 테이블을 조합하여 sales_outlet_id 별 목표 판매 건수와 달성 현황을 알아보았다. 아울렛(3)은 127%로, 예측 판매 목표를 초과 달성 한 것을 볼 수 있고, 아울렛(5)는 목표 달성 실패, 아울렛(8)은 목표 달성으로 결과가 도출됐

다. 해당 도표를 통해 관광거점인 아울렛(5), (8) 보다 거주지역인 아울렛(3) 에서의 달성도가 높다는 점을 눈여겨볼 필요가 있을 것 같다. 해당 브랜드는 관광객보다 현지 사람들에게 접근성이 좋음을 시사하며, 현지를 중심으로 하는 업장과 관광객을 중심으로 하는 업장을 구별하여 전략을 다르게 세울 필요성이 있다. 또한 신규 고객 확보를 위해 현지인 보다, 관광객을 대상으로 한 프로모션 등이 효과가 클 것으로 예상된다.

각 아울렛 별 세부 카테고리를 뜯어보면 다음과 같다.

카테고리 별 Goal 달성도 도출

단위: qty

sales_outlet_id	beans_goal	beans_achieved	RATE	달성여부
3	720	313	43%	X
5	1000	362	36%	X
8	900	493	55%	X

sales_outlet_id	beverage_goal	beverage_achieved	RATE	달성여부
3	13500	20001	148%	O
5	18750	19282	103%	O
8	16875	19822	117%	O

sales_outlet_id	food_goal	food_achieved	RATE	달성여부
3	3420	2489	73%	X
5	4750	2694	57%	X
8	4275	2581	60%	X

sales_outlet_id	merchandise_goal	merchandise_achieved	RATE	달성여부
3	360	94	26%	X
5	500	125	25%	X
8	450	36	8%	X

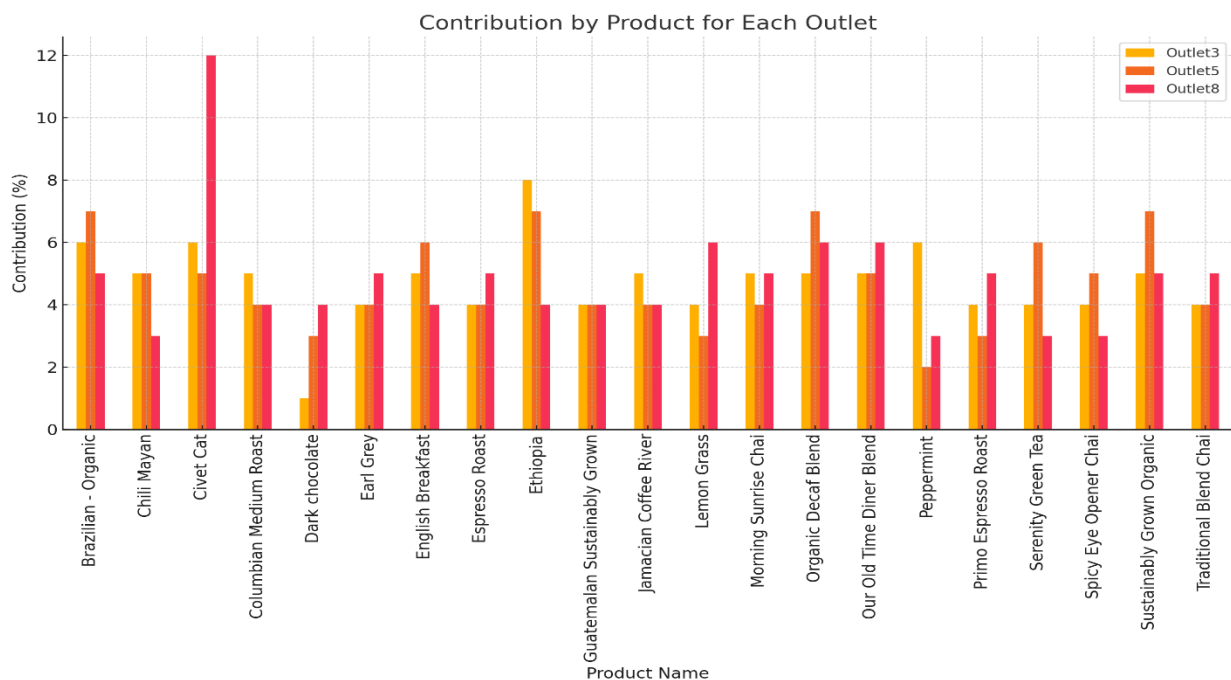
[도표3] 카테고리별 GOAL 달성도

카테고리 구성은 Beans, Beverage, Food, Merchandise 총 4가지로 모든 매장에서 음료 카테고리에 과도하게 의존하고 있음을 확인 할 수 있다.

특히 아울렛(3)은 음료 카테고리에서 148% 달성률을 보인다. 식품 카테고리는 달성률이 미비하나 나머지 카테고리과 비교하면 전 매장 모두 50% 이상은 달성 하였다. 식품 카테고리의 목표 달성가능성이 현실적이라고 판단되며, 제품 다양화 및 시식 행사 등을 통해 관광객의 발길을 멈춰볼 수 있을 것 같다.

그 외 원두와 MD 카테고리는 전 매장 달성률이 50% 넘는 곳이 한 곳만 보인다. 해당 카테고리는 우리나라 브랜드들이 추후 해외 진출시 재고를 예상 실적 대비 적게 운영 하는 등의 전략이 필요해보인다.

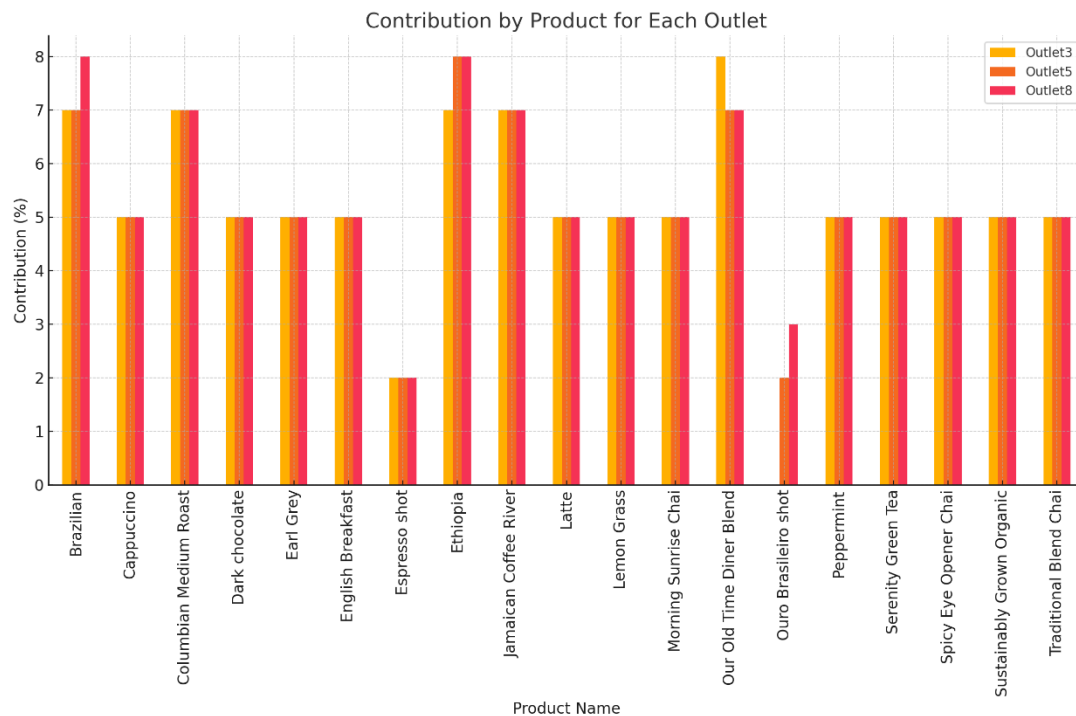
음료를 제외한 타 카테고리서 목표 달성을 하지 못한 원인을 추가적으로 확인해보기 위해 일부 상품의 판매 비중이 현저히 낮았을 것을 가정하고, **각 아울렛당 카테고리 별 상품 종류와 판매량 달성 기여도 비중이** 어떻게 되는지 알아보았다.



[차트2] BEANS 내 판매량 달성 기여도 비교

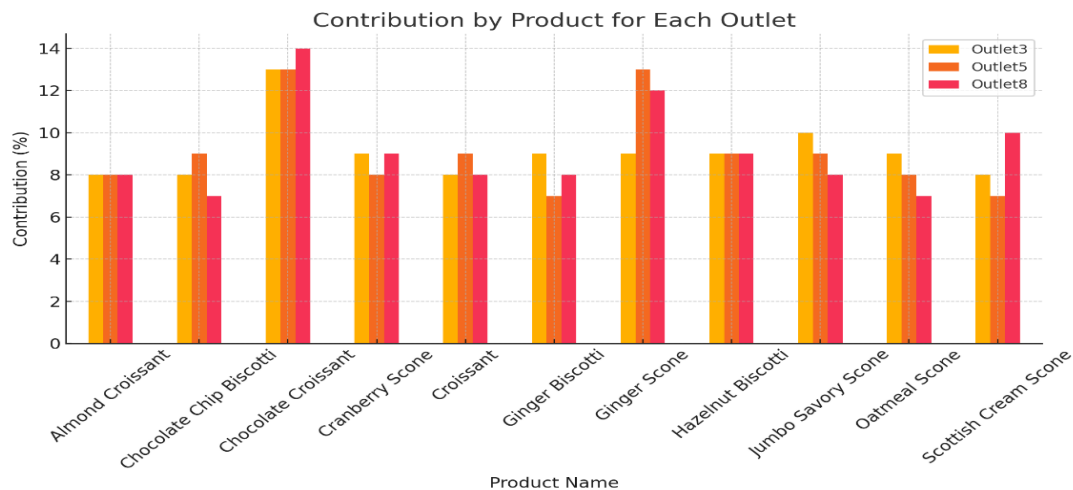
Beans 카테고리는 매장별 달성률이 아울렛(5), 최저 36% 에서 아울렛(8)의 경우 최대 55%로 판매 증진을 위한 추가적인 전략이 필요하다. 카테고리내 품목 별 기여도의 평균을 도출해보면 21개 품목이 모든 매장에서 5%로, 21개의 상품 다 비교적 골고루 판매되

는 걸 볼 수 있다. 'Dark Chocolate' 은 모든 아울렛에서 평균 기여도 이하로 팔리는걸 알 수 있고, 'Civet Cat'의 경우 아울렛(8) 에서 12%을 나타내는데, 한 손님이 집중적으로 구매 한 건지는 추가 확인해볼 필요가 있다.



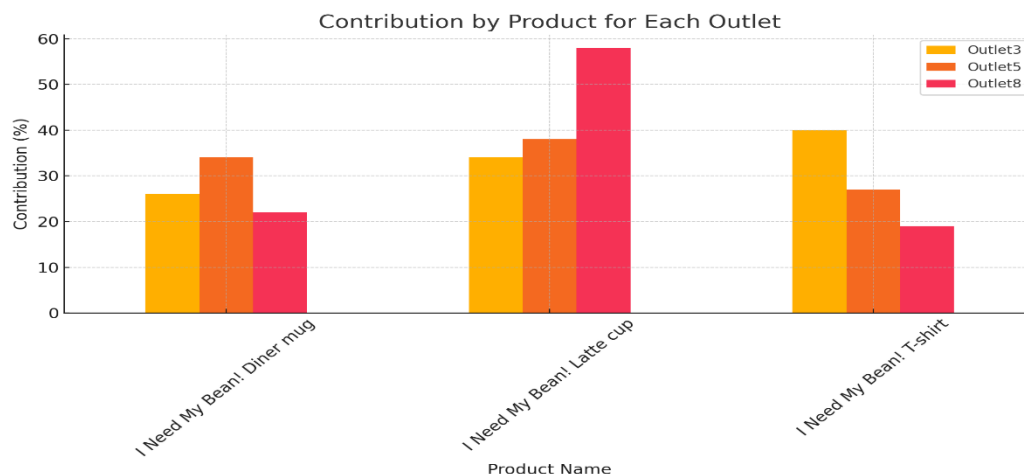
[차트3] BEVERAGES 내 판매량 달성 기여도 비교

Beverages 의 경우, 총 19개 품목으로 'Ouro Brasileiro shot' 의 경우 아울렛(5)와 (8)에서는 프로모션을 진행하는 상품이나, 아울렛(3)에서는 제공하지 않는 걸 확인할 수 있었다. 다만, 이 역시 품목별 평균 기여도가 5%인 걸 고려하면 비중이 낮은 품목이다. 'Beverage' 또한 'Espresso shot' 을 제외하고는 모든 상품이 모든 매장에서 골고루 판매된 걸 확인할 수 있다.



[차트4] FOOD 내 판매량 달성 기여도 비교

모든 아울렛이 50% 이상은 달성한 Food 카테고리다. 총 구성 상품은 9개로, 세 outlet 모두 운영상품이 동일하다. 상품별 평균 기여도는 9%로, 세 아울렛 모두 동일하다. 아울렛 (5)와 (8)에서 프로모션을 진행했던 'Ginger Scone' 은 두 업장에서 평균 이상으로 기여했고, 프로모션의 영향이 있었음을 추측해볼 수 있다. 'Chocolate Chip Biscotti' 는 세 아울렛 모두 12% 이상의 수치로, 주력 상품이였음이 확인 가능하다. 초콜렛 칩을 포함한 상품들로 메뉴 다각화를 고려해보면 좋을 것 같다.



[차트5] MERCHANDISE 내 판매량 달성 기여도 비교

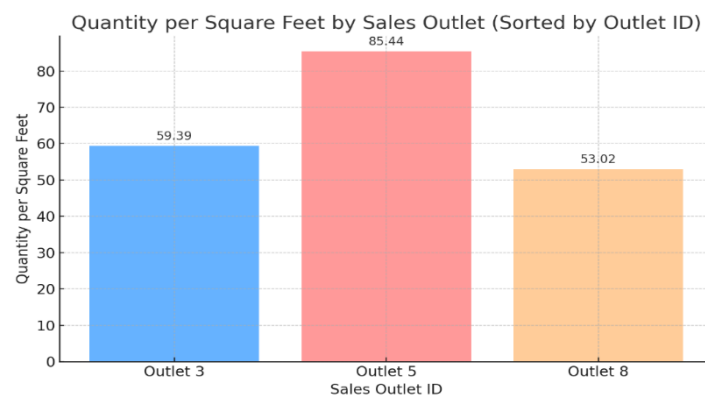
Merchandise 항목은 Mug, Cup, T-shirt 총 3 품목으로 구성되어있고, 모든 아울렛의 최저 달성 카테고리 별다른 패턴이 없어보인다. 판매 달성도가 25% 이거나, 아울렛(8)의

경우 8%로, 해당 카테고리의 재고는 100EA 미만으로 운영하거나, 선오더후생산 방식을 통해 재고 최적화 전략이 필요해보인다.

종합해서 다시 보면, 일부 상품의 판매 비중이 현저히 낮았던 경우는 'Beverages' 내 'Espresso shot' 을 제외하고는 없었으며, 모든 상품이 모든 매장에서 골고루 팔렸던 걸 확인할 수 있었다. 'Food' 의 경우 초콜렛칩이 들어간 상품들로 다양화하여 판매량을 증진하고, 'Merchandise'의 경우 선오더후생산 방식 등 재고 최적화 전략이 필요해 보인다.

4. 그 외 추가 분석

상위 도출 결과를 다양한 각도로 보기 위해 매장 당 평수 대비 발생 매출을 알아보았다.



[차트6] 매장 당 평수 대비 발생 매출 현황

도출 결과, 아울렛(5)이 1평방미터당 85 달러로 평수 대비 가장 높은 매출을 보였다. 아울렛(5)가 유일하게 목표 판매 건수를 달성하지 못한 장소인 점과 유동인구가 가장 많은 지역 또한 아울렛(5)인 점을 고려하면 해당 점포는 목표 건수가 높게 잡혔을 가능성이 있다.

sales_outlet_id	in_store_count	to_go_count
3	8508	8321
5	8010	7984
8	8474	8303

[도표4] OUTLET 별 테이크 아웃 건수 비교

추가로 해당 브랜드의 아울렛 별 매장 취식과 테이크 아웃의 건수를 비교해보았는데, 세 매장 모두 매장 취식과 테이크 아웃의 비율이 50.n : 49.n 로 매장 취식과 포장의 생각이 비슷하다 비슷한 수치를 보였다.

product_group	product_category	product_type	to_go_count
Beverages	Coffee	Gourmet brewed coffee	2876
Beverages	Tea	Brewed Chai tea	2797
Beverages	Coffee	Barista Espresso	2750
Beverages	Tea	Brewed Black tea	1895
Beverages	Drinking Chocolate	Hot chocolate	1863
Beverages	Tea	Brewed herbal tea	1851
Food	Bakery	Scone	1659

[도표5] 상품별 테이크 아웃 건수

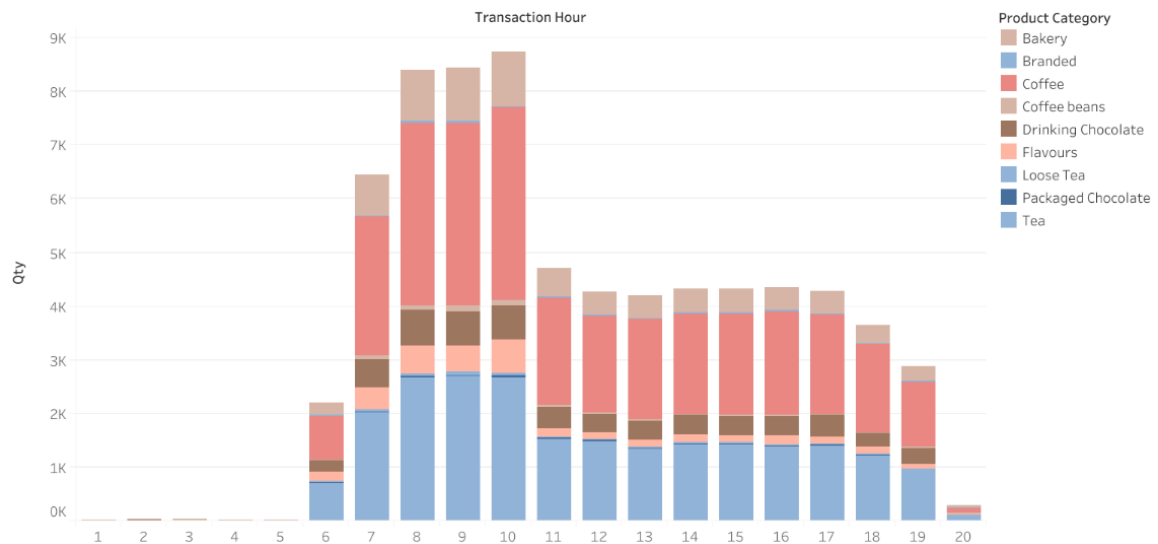
테이크 아웃이 많은 상품들 위주로 조합하여 포장할인행사를 진행하거나, 전 음료 대상 포장시 할인 행사를 진행하여 매장내 분위기를 쾌적하게 유지하고, 유동인구를 대상으로 한 매출증진 전략을 진행해볼 수 있을 것 같다.

주요 상품 데이터 분석(옥예진)

1. 시간대별 상품 판매량 분석

1.1 시간대별 선호 상품

시간대별 선호 상품



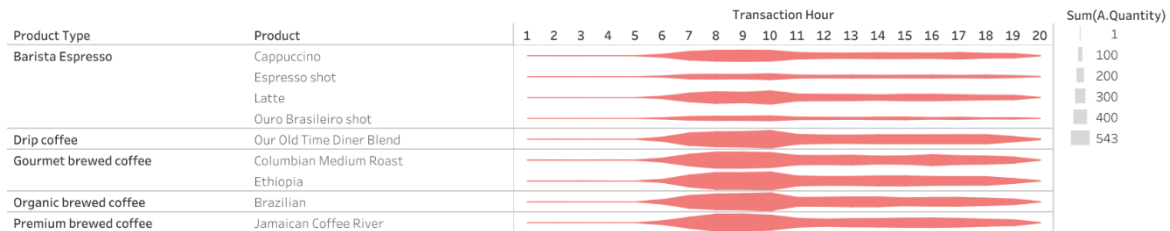
[차트7] 시간대별 선호 상품

- 전체 판매량은 오픈(오전 6 시~7 시) 시간대부터 증가하는 추세를 보이다가 오전 10 시에 가장 높은 판매량을 기록했다. 오전 11 시부터는 판매량이 급격히 감소하면서 오후 시간대의 판매량은 미미한 수준이었다.
- Product Category 기준 모든 시간대의 판매량 1 위는 Coffee 가 기록했으며, 2 위와 3 위는 대부분 Tea, Bakery 순으로 판매량이 높았다.

1.2 시간대별 판매량 세부 추이

- Product Category를 기준으로 한 시간대별 분석에서 더 나아가, 각 Product Category 내에서 시간대별 어떤 Product(즉, 메뉴)가 판매량이 높은지 파악하기 위해 Product 기준 시간대별 판매량 세부 추이를 분석해보았다.
- 이때, 데이터 내 Product가 Lg, Rg, Sm의 Size로 구분되어 있어, Product별 정확한 판매량을 파악하기 위해 Size를 통합하는 데이터 전처리 후 분석을 진행했다.

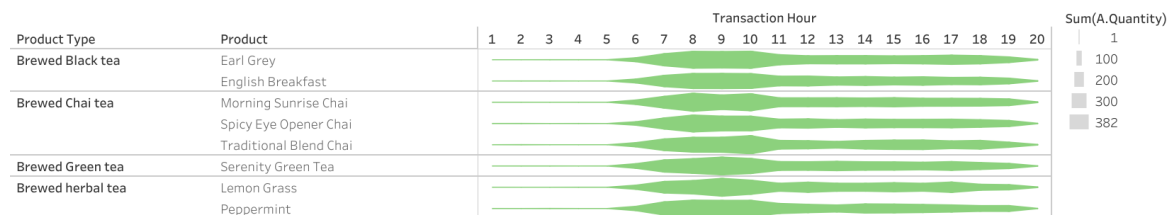
1.2.1 시간대별 Coffee 판매 추이



[차트8] 시간대별 COFFEE 판매 추이

- 오전 시간대 최고 판매량(16 시)을 기록한 Product 는 'Gourmet brewed coffee' Type 의 'Ethiopia'였으며, 오후 시간대의 최고 판매량(16 시)을 기록한 Product 또한 'Gourmet brewed coffee' Type 의 'Columbian Medium Roast'였다.
- 커피는 주로 아침 식사와 함께 마시는 미국의 커피 문화로 인해 오전 판매량이 오후 판매량보다 압도적으로 많은 것으로 보인다.

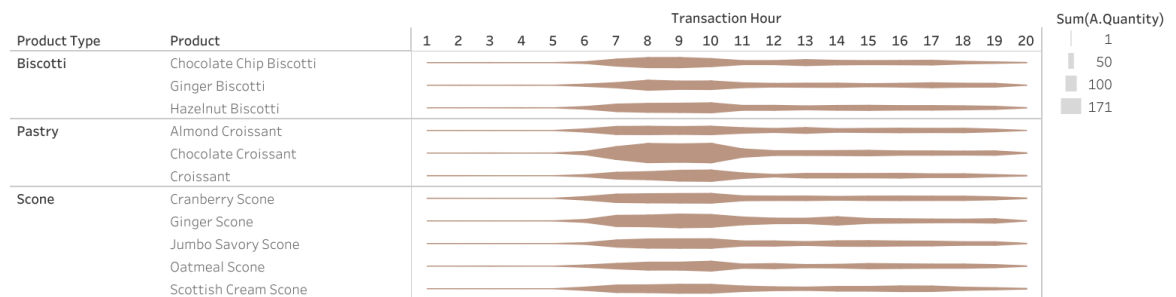
1.2.2 시간대별 Tea 판매 추이



[차트9] 시간대별 TEA 판매 추이

- 오전 시간대 최고 판매량(10시)을 기록한 Product는 'Brewed Chai tea' Type의 'Traditional Blend Chai'였으며, 오후 시간대 최고 판매량(17시)을 기록한 Product는 'Brewed herbal tea' Type의 'Lemon Grass'였다.

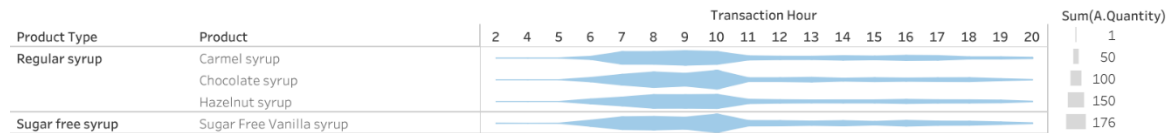
1.2.3 시간대별 Bakery 판매 추이



[차트10] 시간대별 BAKERY 판매 추이

- 오전 시간대 최고 판매량(10시)을 기록한 Product는 'Pastry' Type의 'Chocolate Croissant' 였으며, 오후 시간대 최고 판매량(14시)을 기록한 Product는 'Scone' Type의 'Ginger Scone' 이었다.
- 'Ginger Scone'의 높은 판매량은 프로모션의 변수를 고려해야 한다.

1.2.4 시간대별 Flavours 판매 추이

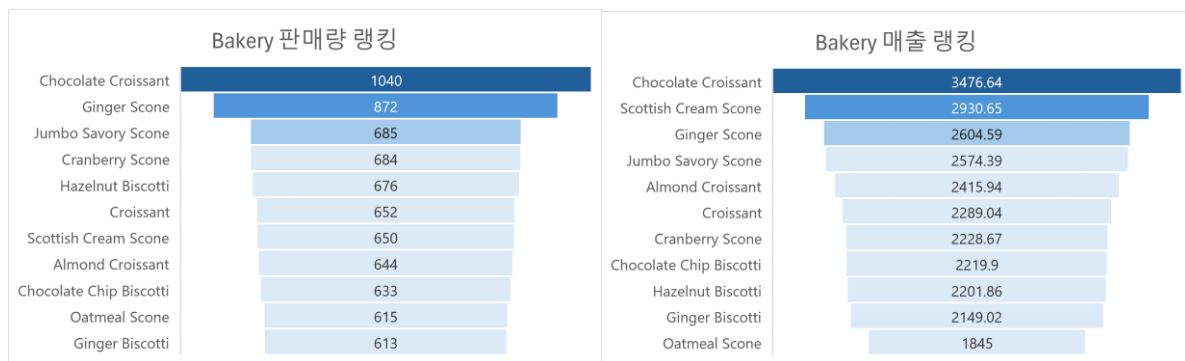


[차트 11] 시간대별 FLAVOURS 판매 추이

- 오전 시간대 최고 판매량(10시)을 기록한 Product는 'Regular Syrup' Type의 'Chocolate Syrup' 였으며, 오후 시간대 최고 판매량(16시)을 기록한 Product는 'Sugar Free Syrup' Type의 'Sugar Free Vanilla syrup' 였다.
- 오전과 오후 시간대 모두 'Sugar Free Syrup'에 대한 선호도가 꾸준히 높은 것으로 나타났다.

2. Bakery 데이터 분석

2.1 Bakery 판매량 및 매출 랭킹

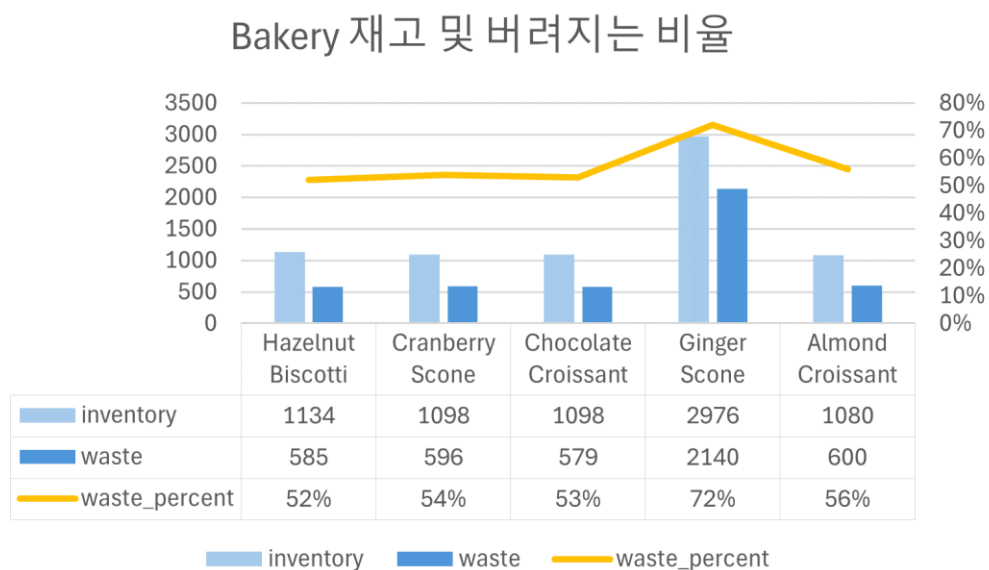


[차트12] BAKERY 판매량 및 매출 랭킹

- 'Chocolate Croissant'은 판매량과 매출 모두에서 가장 높은 수치를 기록하고 있으므로, 인기가 많은 상품으로 주력해볼 수 있다.
- 'Oatmeal Scone'은 판매량과 매출 모두에서 낮은 수치를 기록하고 있다.
- 'Ginger Scone'의 높은 판매량 및 매출은 프로모션의 변수를 고려해야 한다.

2.2 Bakery 재고

2.2.1 Bakery 재고 및 버려지는 비율



Inventory: 재고 수 / waste: 버려지는 수 / waste_percent: 버려지는 비율

[차트13] BAKERY 재고 및 버려지는 비율

- Bakery 재고 데이터 중 재고 수와 버려지는 수의 비율을 계산해보니, 데이터에 존재하는 모든 Bakery가 버려지는 비율이 50% 이상으로 나타났다.

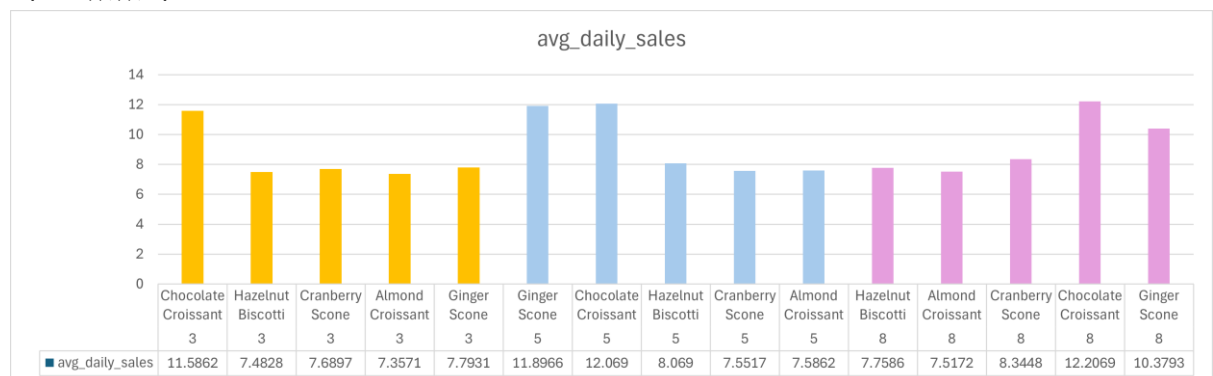
- 특히 Ginger Scone은 프로모션을 했음에도 버려지는 비율이 70% 이상으로 확인되므로 인기가 높은 상품은 아님을 시사한다.

2.2.2 Bakery 수요 예측 오류 분석

transaction_date	product	inventory
4/1/2019	Hazelnut Biscotti	18
4/1/2019	Cranberry Scone	18
4/1/2019	Chocolate Croissant	18
4/1/2019	Ginger Scone	48
4/1/2019	Almond Croissant	18

[도표6] BAKERY 별 일일 운영 재고 수량

- 모든 지점에서 매일마다 'Ginger Scone'은 48 개, 그 외 4 개의 Bakery 는 18 개의 재고를 일정하게 채워두고 있었고, Bakery 특성상 팔리지 않은 상품은 당일 폐기를 하고 있었다.

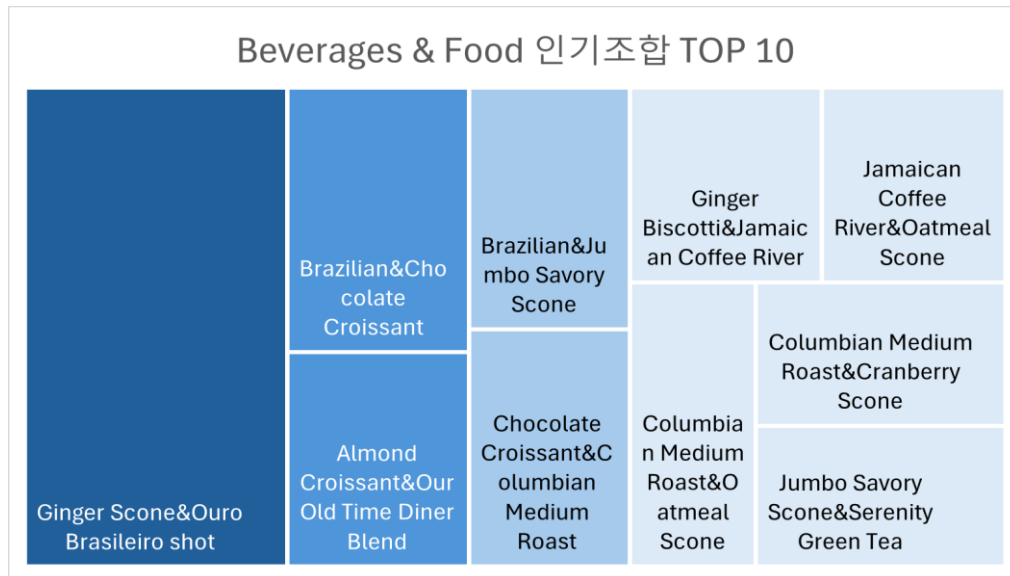


[차트 14] 지점별 BAKERY 일일 평균 판매량

- 지점마다 Bakery의 일일 평균 판매량을 계산해봤을 때, 대부분의 판매량이 10개를 웃돌았다. 이때, 'Ginger Scone'의 경우 매일마다 48개의 재고를 확보해왔지만, 실제 판매량은 재고의 4분의 1도 되지 않았다. 이는 수요에 비해 과도한 공급이 이루어졌음을 시사한다.
- 따라서 많은 재고가 버려지고 있었던 원인은 수요 예측에서 오류가 발생했기 때문이라고 유추할 수 있다. 유통기한이 짧아 당일 폐기를 해야 하는 Bakery의 특성상, 더욱 정확한 수요 예측에 따라 재고를 조율할 필요가 있음을 시사한다.

3. 인기 조합 분석

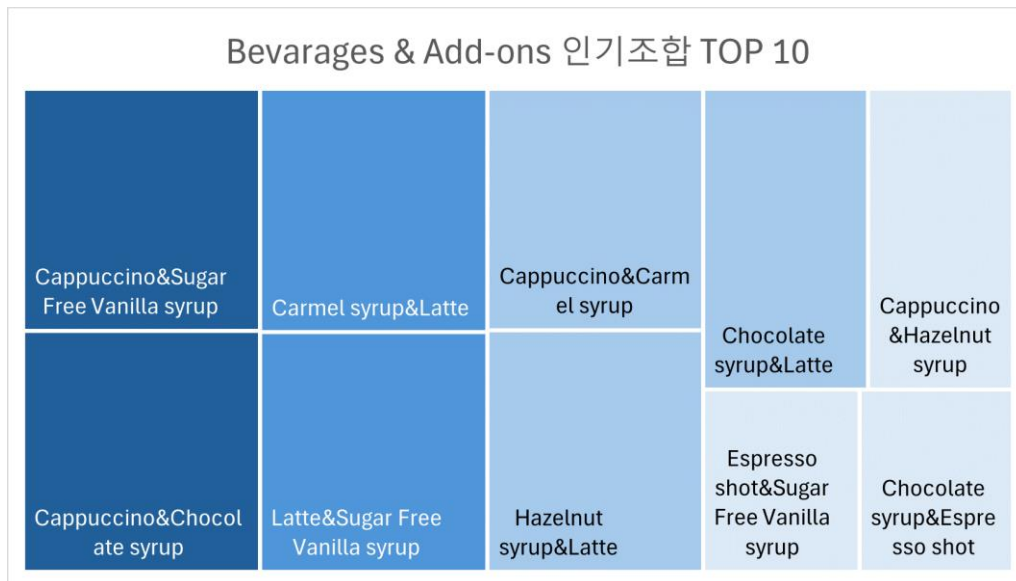
3.1 음료 & 베이커리 조합



[차트15] BEVERAGES & FOOD 인기조합

- 고객들이 가장 많이 구매한 음료와 베이커리의 조합으로는 Scone과 Espresso의 조합인 'Ginger Scone' & 'Ouro Brasileiro shot' 이었다. 하지만, 이는 프로모션의 영향으로 추측되어 뒤이어 '4. 프로모션 효과 분석'에서 더 자세히 살펴보겠다.
- 위의 그래프에서 3번의 빈도수로 나타나, 다른 음료 상품보다 가장 자주 베이커리와 함께 구매한 음료는 'Columbian Medium Roast'였다. 부드러움이 특징인 해당 커피와 함께 베이커리를 자주 구매하는 경향은 부드러운 커피와 베이커리의 조합이 선호도가 높음을 시사한다.

3.2 음료 & 시럽 조합



[차트16] BEVARAGES & ADD-ONS 인기조합

- 고객들이 가장 많이 구매한 음료와 시럽의 조합으로는 'Cappuccino' & 'Sugar Free Vanilla syrup' 이었다. 이는 건강관리에 관심이 높아진 트렌드에 따라 고객들의 저당 시럽에 대한 높은 선호도를 시사한다.
- 'Latte'에 'Caramel syrup', 'Hazelnut syrup', 'Chocolate syrup'을 넣어 먹는 조합도 많은 것으로 나타났다.

4. 프로모션 효과 분석

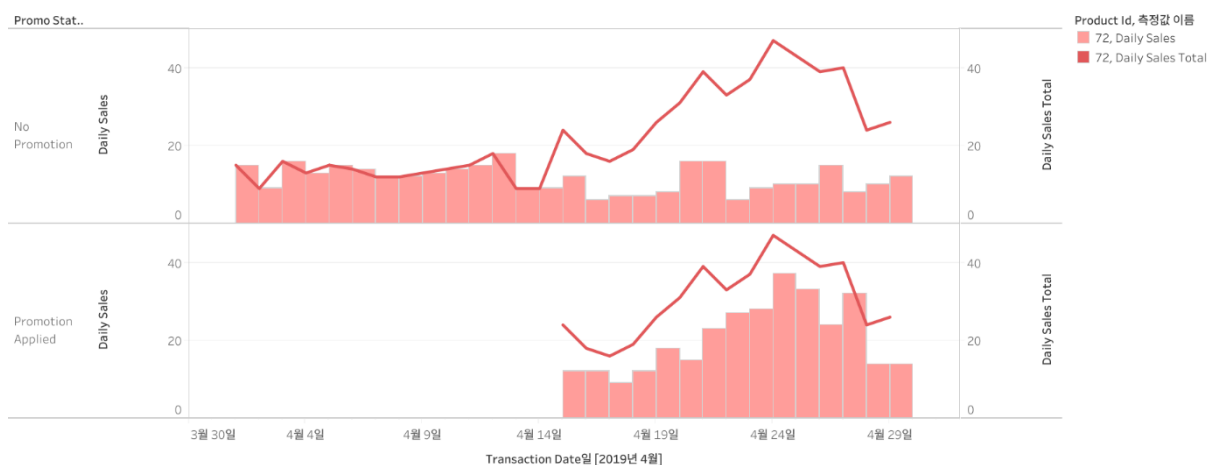
4.1 프로모션 효과 분석 배경

- 본 데이터에는 4월 15일~29일(총 15일) 동안 sales_outlet_id가 5, 8인 곳에서 'Ginger Scone'과 'Ouro Brasileiro shot'의 프로모션을 진행한 기록이 남아있다.
- 앞서 '2.1 Bakery 판매량 및 매출 분석'과 '3.1 음료 & 베이커리 조합'에서 나타난 'Ginger Scone'의 높은 판매량 및 매출 기록과 'Ginger Scone' & Ouro Brasileiro shot' 조합의 높은 판매량이 프로모션의 영향인지 아닌지를 분석하기 위해 프로모션 효과 분석을 진행했다.

4.2 가설 설정 및 분석 방안

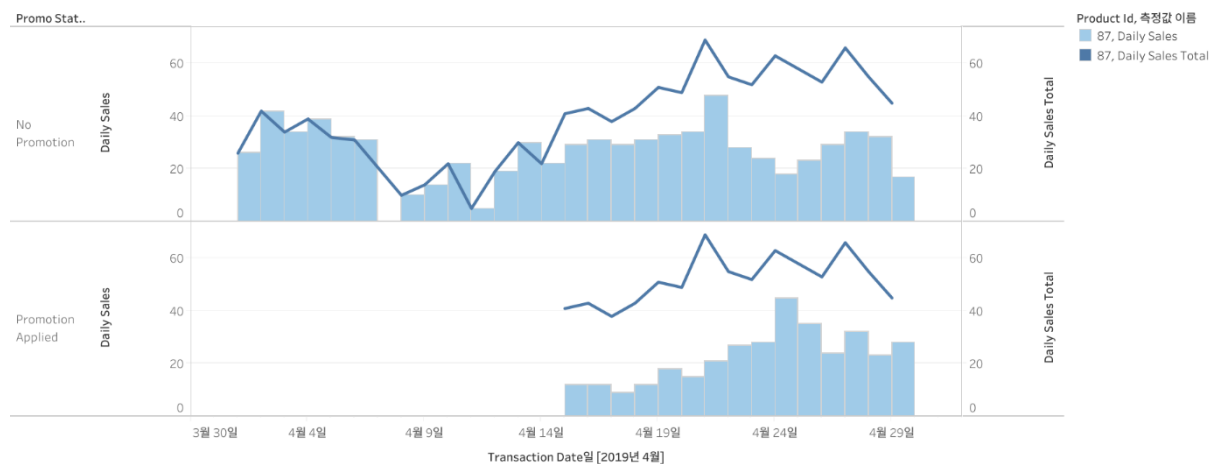
- 프로모션을 하면 판매량이 증가할 것이라는 가설을 설정했고, 더하여 'Ginger Scone'과 'Ouro Brasileiro shot'의 높은 판매량은 프로모션의 영향이었음을 가정했다.
- Product_id별로 프로모션 적용 여부에 따라 나누어 일별 판매량 변화를 확인했다.
- 프로모션은 sales_outlet_id가 5, 8인 지점에서만 진행되고 있었으므로, 보다 정확한 판단을 위해 프로모션을 하지 않은 sales_outlet_id가 3인 지점은 배제 후 5, 8인 지점의 판매량만 합쳐서 계산했다.
- 프로모션 기간에 프로모션 가격으로 판매된 건 'Promotion Applied'로 분류했고, 프로모션 기간 이전이거나, 프로모션 기간 내 프로모션 상품임에도 프로모션 가격이 적용되지 않고 판매됐던 내역은 'No Promotion'으로 분류했다.
- 시각화 과정에서 프로모션 여부에 따른 일일 판매량(Daily Sales)은 막대 그래프로, 프로모션 여부를 적용하지 않은 일일 총 판매량(Daily Sales Total)은 라인 그래프로(이중축)로 표시했다.

4.3 분석 결과



[차트17] GINGER SCONE 프로모션기간 전후 판매량 비교

- 'Ginger Scone'(Product_id가 72인 상품)은 비프로모션 기간 일일 평균 판매량이 13.1개였던 반면, 프로모션 기간에는 약 30.8개로, 135.1145% 증가했다.
- 라인 그래프(프로모션 여부를 적용하지 않은 일일 총 판매량)를 확인해보았을 때, 프로모션 기간부터 라인 그래프가 상승선을 그리고 있음을 알 수 있다.
- 이에 따라 'Ginger Scone'의 판매량 증가는 프로모션의 영향이 있었다고 판단할 수 있다.



[차트18] ORO BRASILEIRO SHOT 프로모션기간 전후 판매량 비교

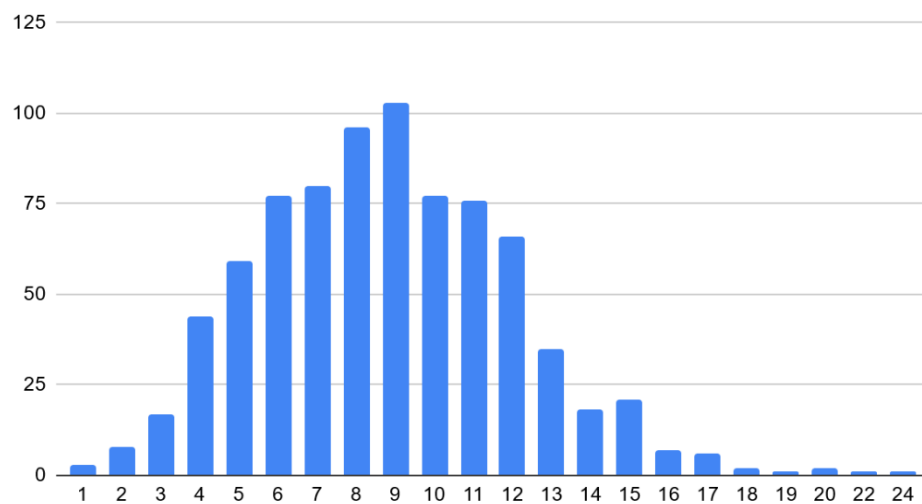
- 'Ouro Brasileiro shot'(Product_id가 87인 상품)의 비프로모션 기간 일일 평균 판매량은 25개였던 반면, 프로모션 기간에는 약 52개로, 108% 증가했다.
- 라인 그래프(프로모션 여부를 적용하지 않은 일일 총 판매량)를 확인해보았을 때, 프로모션 기간부터 라인 그래프가 상승선을 그리고 있음을 알 수 있다.
- 이에 따라 'Ouro Brasileiro shot'의 판매량 증가는 프로모션의 영향이 있었다고 판단할 수 있다.
- 결론적으로, 'Ginger Scone'과 'Ouro Brasileiro shot' 모두 프로모션을 한 이후부터 판매량이 100% 이상 증가했으므로 프로모션의 효과가 있었다고 분석할 수 있다. 'Ginger Scone'과 'Ouro Brasileiro shot'의 판매량이 높은 수치를 기록했던 것 또한 프로모션의 영향이었음을 알 수 있다.
- 한편, 개별적으로 프로모션되었던 'Ginger Scone'과 'Ouro Brasileiro shot'을 함께 구매하는 경향이 높았던 점에도 주목해볼 수 있다. 각 제품을 개별적으로 프로모션 할 수도 있지만, 함께 구매하도록 유도하는 번들 프로모션을 활용하는 방안도 효과적일 수 있음을 시사한다.

재방문 데이터 분석(황채원)

1. 방문 횟수 분석

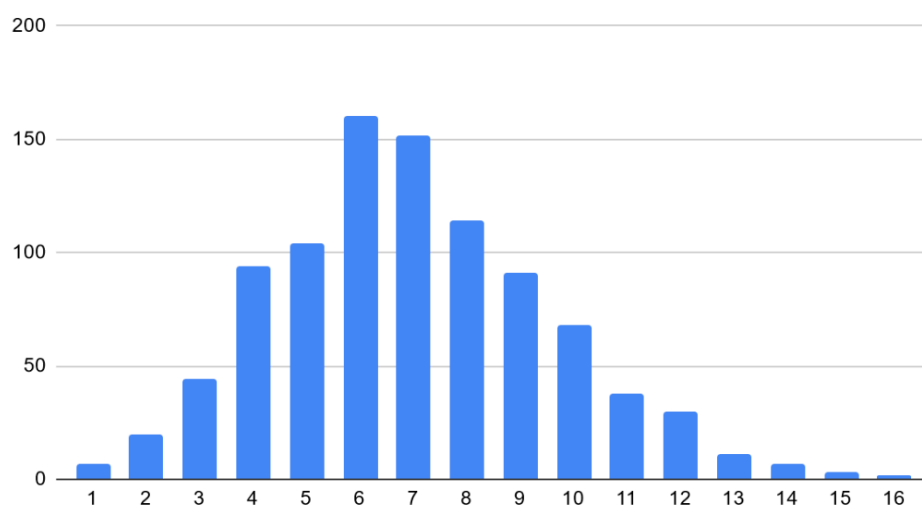
1.1 방문 횟수 집계

outlet3의 방문횟수 별 고객 수



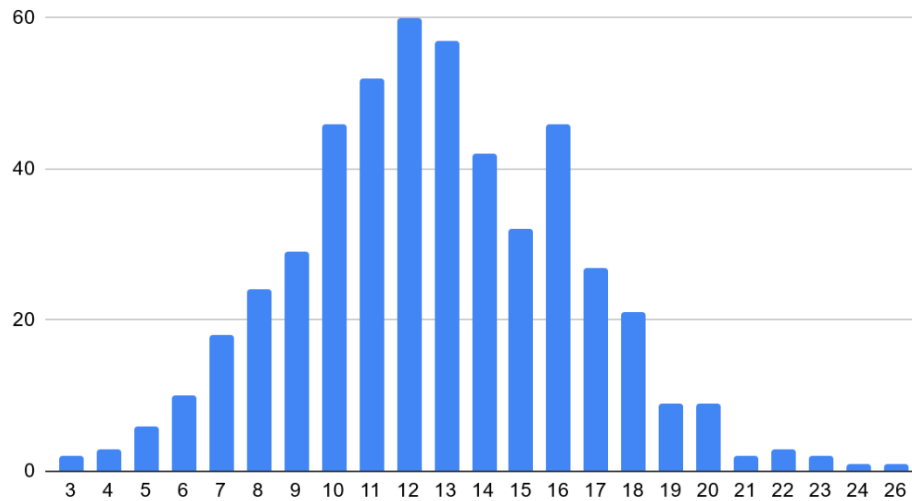
[차트19] OUTLET3의 방문횟수 별 고객 수

outlet5의 방문횟수 별 고객 수



[차트20] OUTLET5의 방문횟수 별 고객수

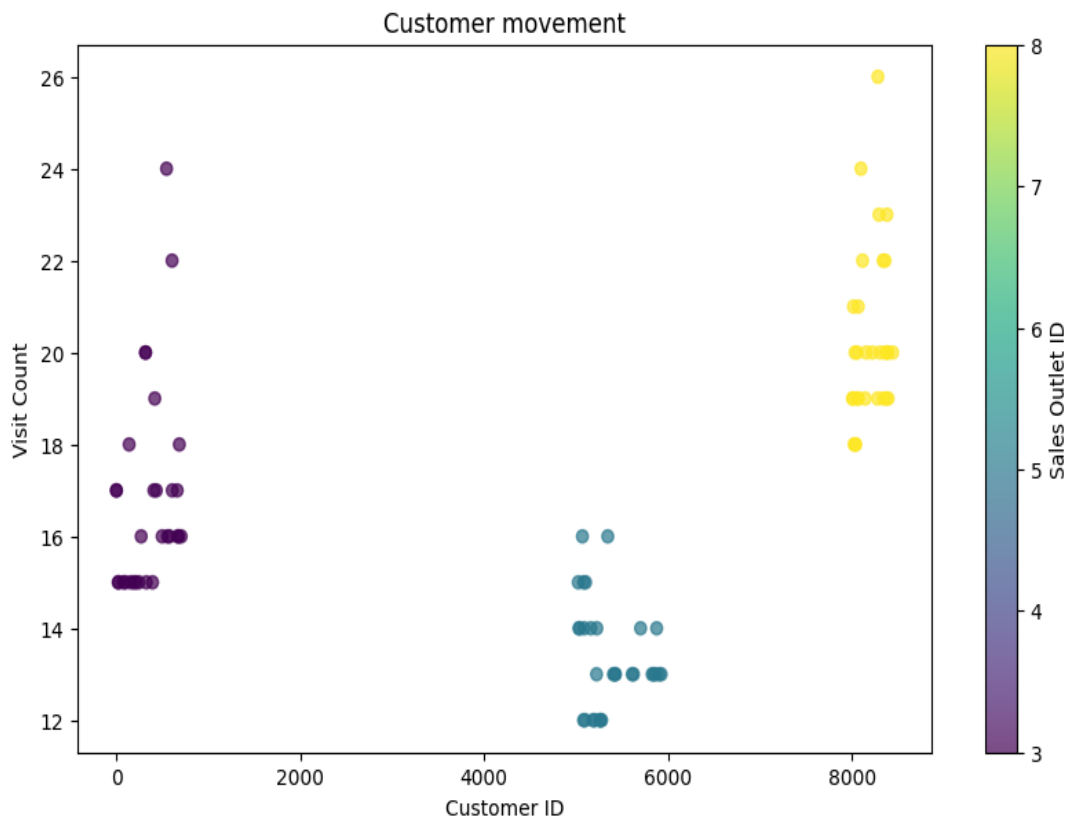
outlet8의 방문횟수 별 고객 수



[차트21] OUTLET 8 방문횟수 별 고객 수

- 각 매장의 고객의 반복 방문 횟수를 집계한 결과, 한 달에 2회 이상 방문한 고객의 수의 비중이 매우 높음을 알 수 있다.
- 그 중에서도 6~11 회씩 방문한 고객의 수는 약 450명으로, 이를 통해 한 달에 같은 가게를 자주 찾는 고객, 즉 충성고객의 존재를 확인할 수 있었다.

1.2 매장별 방문 고객의 id 확인



[차트22] 매장별 방문고객 ID 분포

- 위 산점도를 보면 각 outlet_id 별 customer_id 가 겹치지 않음을 알 수 있다.
- 이를 통해 한 customer는 한 outlet만 이용한다는 것을 유추할 수 있다.

1.3 고객데이터 중복 확인

customer_id	home_store	customer_first-name	customer_email	customer_since	loyalty_card_number	birthdate	gender	birth_year
598	3	Laurel Rutledge	Porter@Pellentesque.gov	2018-09-10	045-914-3248	1988-04-03	F	1988
5437	5	Zelda	Porter@pellentesque.gov	2019-04-06	413-974-4744	2001-03-17	N	2001

[도표7] 고객데이터 중복 여부 검증

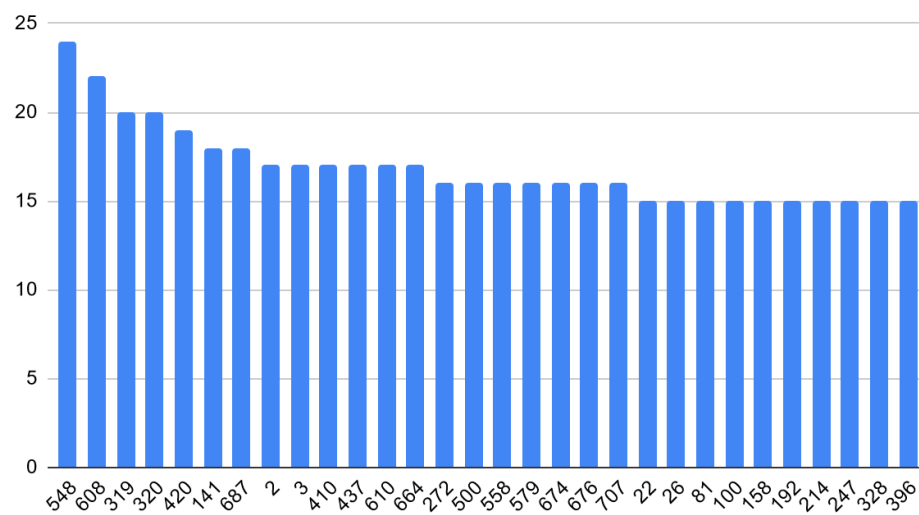
- 1.2의 내용을 검증하기 위해 customer 데이터에서 customer_id는 다르되, 그 외 정보가 같은 사람이 있는지를 확인했다.

- 위 표는 그 결과이며, 이를 통해 customer 데이터에는 같은 사람의 정보가 서로 다른customer_id로 저장되어 있는 경우는 없다고 확신할 수 있다.
- 1.2의 내용과 더불어, 한 customer는 한 outlet만 이용한다는 것을 알 수 있다.

2. 매장별 충성 고객 분석

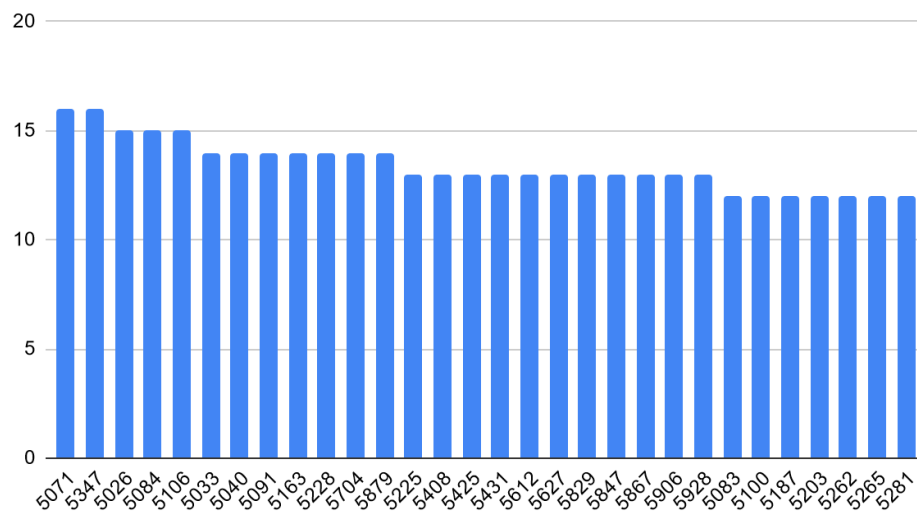
2.1 매장별 고객 행동

outlet3의 고객의 반복 방문 횟수



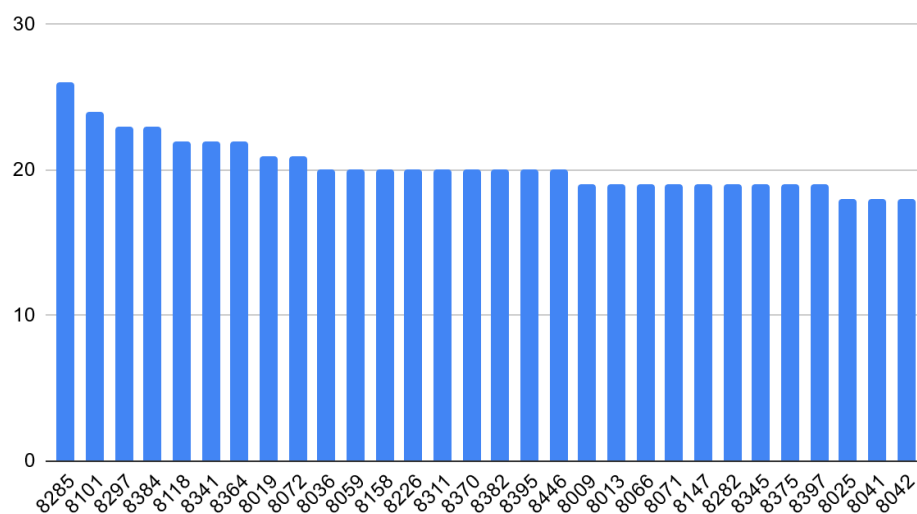
[차트23] OUTLET3의 고객별 반복 방문 횟수

outlet5의 고객의 반복 방문 횟수



[차트24] OUTLET5의 고객별 반복 방문 횟수

outlet8의 고객의 반복 방문 횟수

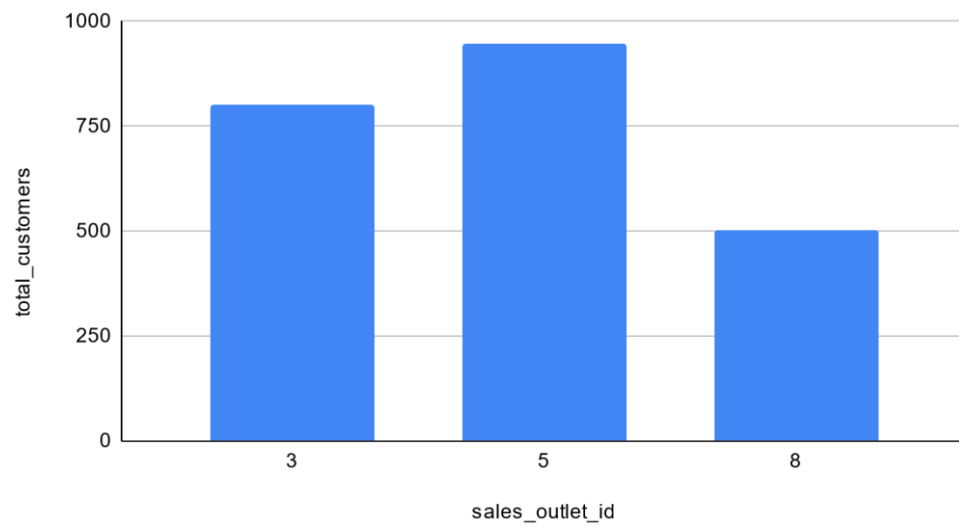


[차트25] OUTLET8의 고객별 반복 방문 횟수

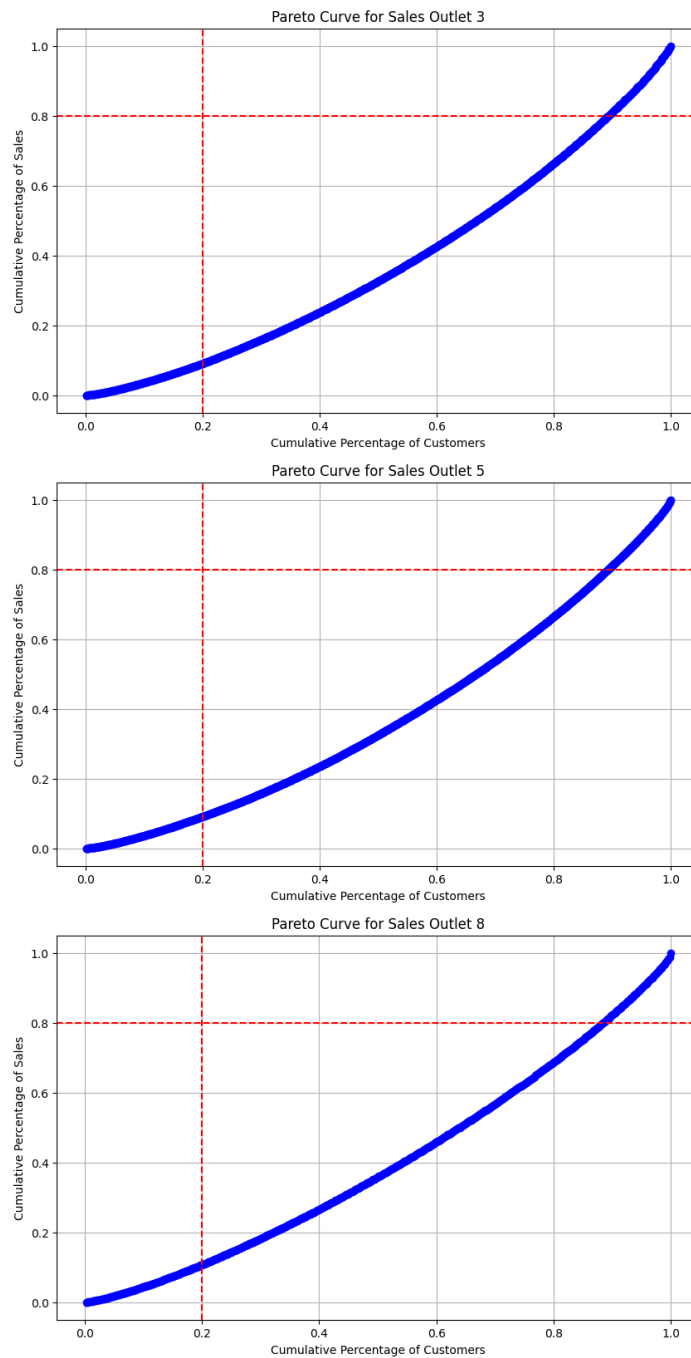
- 각 매장별로 단골 고객이 방문한 횟수를 집계해보면, 최대 1일 1회에 근사한 빈도로 방문하는 고객들이 있다.
- 이는 고객별로 각 매장에 높은 선호도를 보이고 있음을 나타내고, 해당 고객들이 각 매장의 매출에 높은 비중을 차지하고 있음을 유추할 수 있다.

2.2 매출비중 분석

각 outlet 별 방문 고객의 수



[차트26] OUTLET 별 방문 고객 수



[차트27] OUTLET별 상위 방문횟수 고객의 발생 매출 비중

- 각 매장별로 한 달에 방문하는 총 고객의 수는 최소 500명 이상이다.
- 각 매장별로 고객을 방문횟수의 순서대로 정렬한 뒤, 누적 매출을 구한 다음 파레토 곡선을 그린 결과 세 매장 모두 방문횟수 상위 20%의 고객의 매출이 전체 매출의 40% 가까이 된다는 것을 확인할 수 있다.

- 고로 방문 횟수가 낮은 일회성 고객 및 비 단골 고객보다는 방문을 자주하는 단골 고객들의 주요 구매 품목을 파악하는 것이 매출 상승 및 유지에 중요하다.

2.3 충성고객의 구매 패턴

sales_outlet_id	product	total_quantity
3	Earl Grey Rg	288
3	Dark chocolate Lg	283
3	Peppermint Rg	281
3	Sustainably Grown Organic Lg	275
5	Jamaican Coffee River Lg	277
5	Ethiopia Rg	251
5	Latte Rg	248
5	Cappuccino Lg	246
8	Ouro Brasileiro shot	336
8	Ethiopia Lg	276
8	English Breakfast Lg	275
8	Brazilian Lg	274

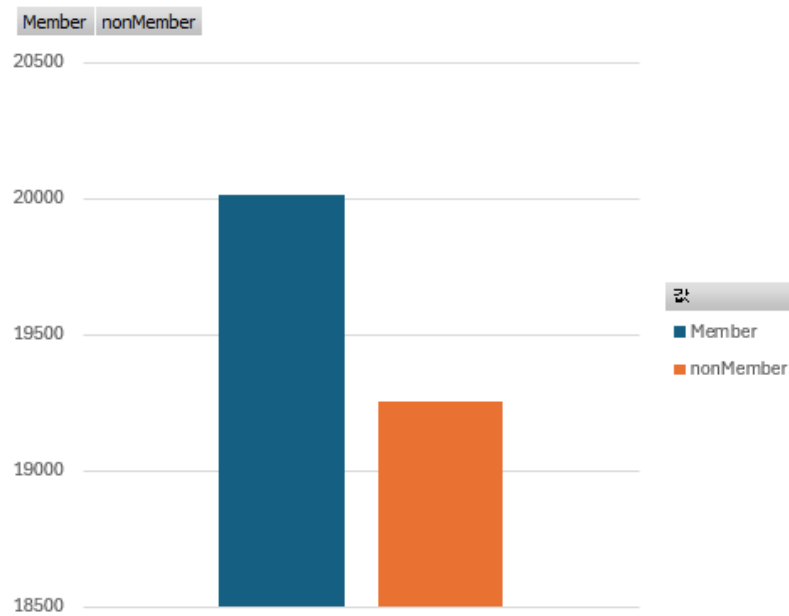
[도표8] OUTLET별 충성고객의 구매 상품 및 수량 집계

- 구매기록 데이터에서 메인 메뉴인 Coffee나 Bakery에 시럽과 같은 부가적인 재료를 추가하는 경우에도 transaction이 중복으로 기록되어 있었기 때문에 이를 포함하는 Flavours 구매기록을 제외하고 각 outlet 별 단골 고객 30명의 메인 메뉴 구매 기록을 집계했다.
- 각 매장별로 단골 고객들이 많이 구매한 메인 메뉴 품목을 4개까지 집계한 결과, outlet3 에서는 'Earl Grey Rg', 'Dark chocolate Lg', 'Peppermint Rg', 'Sustainably Grown Organic Lg' 의 판매량이 많았고, outlet5 에서는 'Jamaican Coffee River Lg', 'Ethiopia Rg', 'Latte Rg', 'Cappuccino Lg'의 판매량이 많았고, outlet8 에서는 'Ouro Brasileiro shot', 'Ethiopia Lg', 'English Breakfast Lg', 'Brazilian Lg' 의 판매량이 많았다.
- 따라서 같은 지역에 국내브랜드의 매장을 새로 개설할 경우 위와 같은 품목들을 중심으로 메뉴를 구성하고 프로모션을 진행하는 것이 매출 상승 및 유지에 효과적일 것이다.

회원 데이터 분석(조수진)

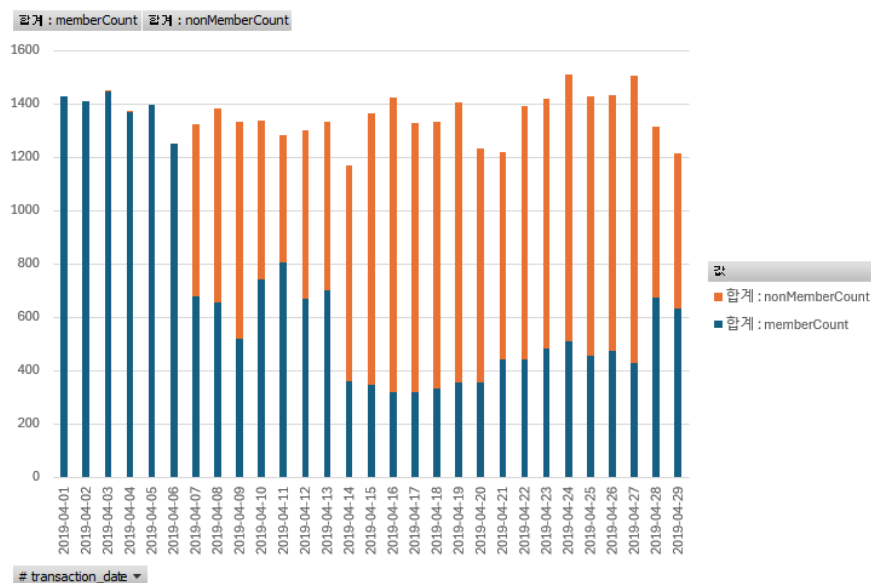
1. 회원 데이터 분석 개요

- 우선 전반적으로 매출 데이터를 분석해 보았을 때, 회원 구매 건수가 비회원 구매 건수보다 약 800 건 높았다.



[차트28] 회원여부에 따른 구매 건수

- 그러나, 일자별로 비회원/회원 매출 추이를 보았을 때, 비회원 매출의 비중이 높아지는 것을 확인할 수 있었다.



[차트29] 거래 일자별 회원/비회원 비중

- 이러한 현상이 발생하는 이유에 대해, 회원 혜택에 대한 메리트가 없기 때문이라고 가설을 생성하고 검증을 진행하였다. 우선, 회원들이 회원 혜택을 받은 적이 있는지 매출 데이터 상에서 확인을 진행하였다. 그리고 두번째로는, 매출을 발생 시킨 회원들이 최근에 가입을 한 회원들인지 확인을 진행하였다.

2. 회원 데이터에 대한 가설 검증

2.1: 회원 혜택 존재 여부 확인

2.1.1 회원 할인 적용 여부 확인

- Promo_item_yn이 'Y'인 상품에 대해 조회하면, 비회원들도 할인된 가격에 음료를 구매한 것을 확인할 수 있었다.

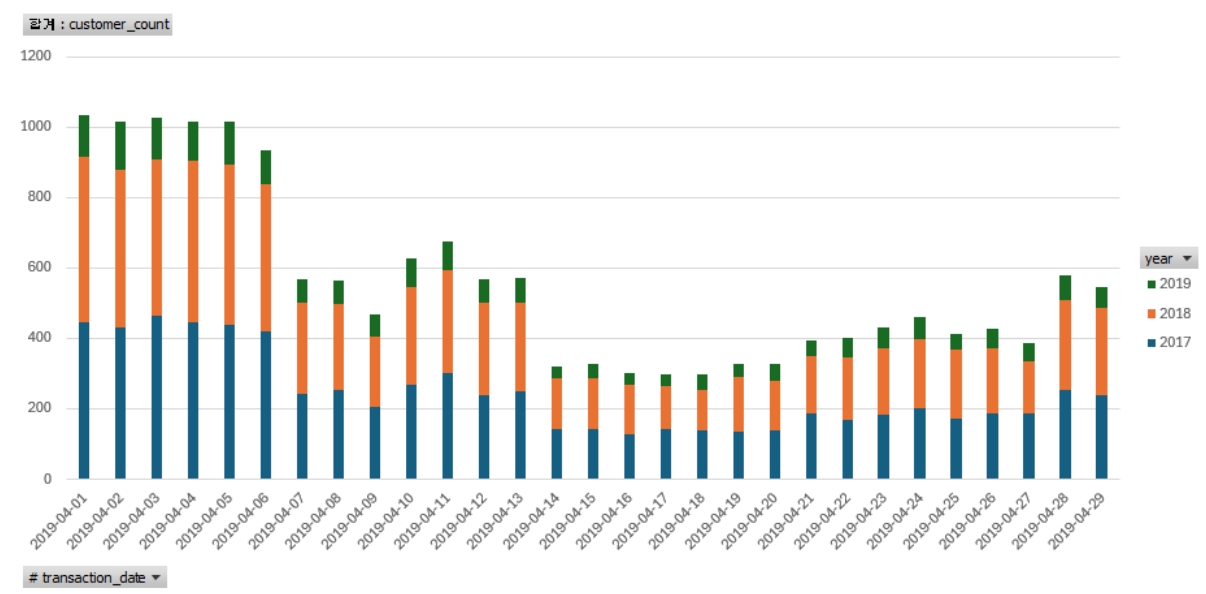
	product_id	product	transaction_date	promo_item_yn	customer_id	unit_price	current_retail_price
회원	87	Ouro Brasileiro shot	2019-04-28	Y	0	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-28	Y	8387	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-28	Y	8149	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-29	Y	5794	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-29	Y	5240	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-29	Y	5109	2.1	\$3.00
비회원	87	Ouro Brasileiro shot	2019-04-29	Y	5662	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-29	Y	5034	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-29	Y	5169	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-29	Y	5670	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-29	Y	0	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-29	Y	8082	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-29	Y	8107	2.1	\$3.00
	87	Ouro Brasileiro shot	2019-04-29	Y	8401	2.1	\$3.00

[도표9] 회원 할인 적용 여부 검증

2.1.2 무료 음료 혜택 확인

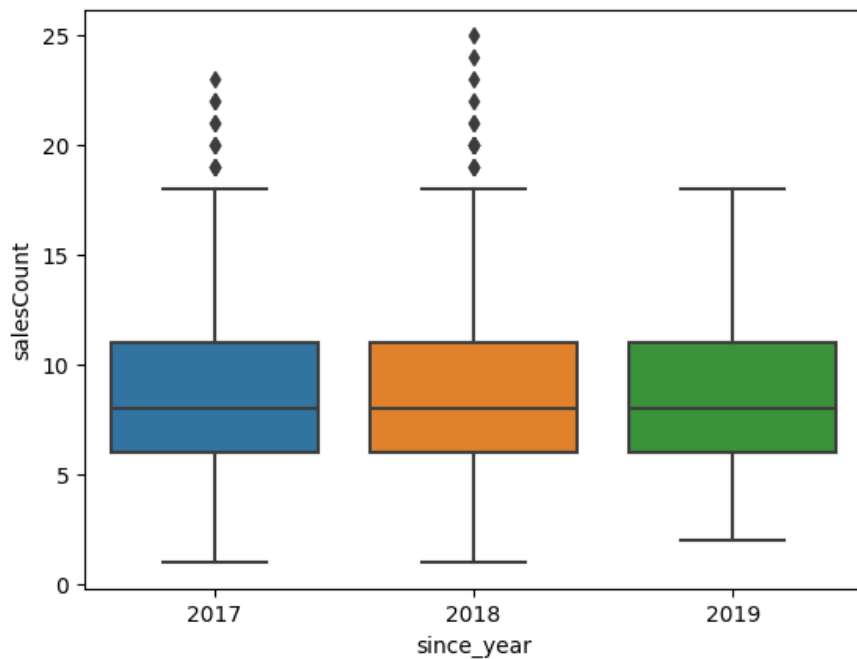
- 쿠폰 사용 여부에 대해 알 수 있는 컬럼이 없어, unit_price가 0인 경우에 대해 조회를 진행하였는데, 조회가 되는 데이터가 없었다.

2. 매출을 발생 시킨 회원들이 최근에 가입을 한 회원들인지에 대한 확인



[차트30] 구매일자 별 이용 회원 가입년도 비교

- 해당 데이터를 통해, 2017년~2018년에 가입한 고객들이 주로 카페를 이용한다는 것을 알 수 있었다.
- 조금 더 자세히 알아보기 위해, 회원들이 매출 데이터 집계 기간 내에 음료를 얼마나 자주 사마셨는지에 대해 확인을 진행하였다.



[차트31] 회원 가입년도 별 음료 소비 분포

- 전체적인 분포 확인을 위하여 boxPlot을 이용하여 시각화를 하였다. 대부분의 고객들은 매출 데이터 집계 기간 내에 6회~11회 정도 음료를 사마셨다. 그러나, 2017년도~2018년도에 가입한 고객들 중에서는 약 18회 이상 음료를 사마신 고객들이 있었다. 이러한 분석을 기반으로, 18회 이상 방문한 고객들을 단골 고객, 6~11회 이상 방문한 고객들을 비 단골 고객이라고 정의하였다.

가입년도	전체 회원 수	한달 평균 회원 증가 수	단골 고객 수	차지하는 비중
2017년	985	82	20	2.03%
2018년	988	168	27	2.73%
2019년	272	68	7	2.57%

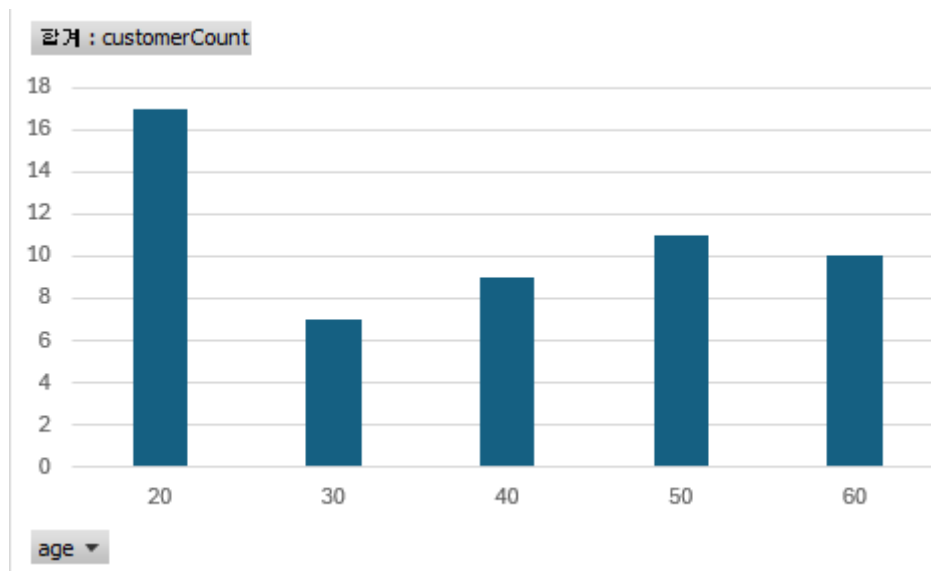
[도표10] 가입년도 별 단골 고객 집계

- 가입년도 별 전체 회원 수를 집계해 보았을 때, 2019년도의 회원수가 다른 년도

에 비해 낮았다.

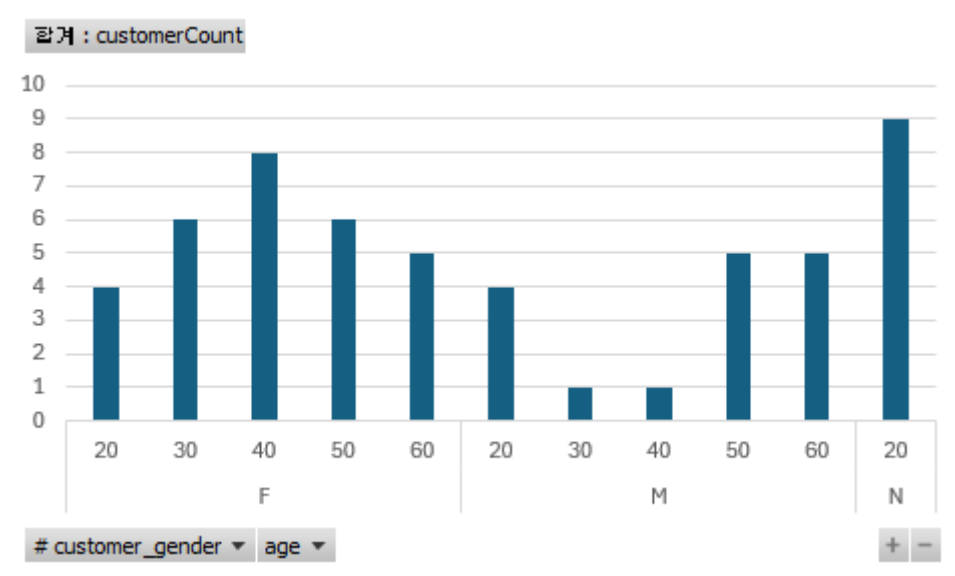
- 집계된 매출 데이터가 2019년도 4월 데이터 임을 감안하면, 회원수가 낮지 않다고 볼 수도 있었을 것이다. 그러나, 2017년도, 2018년도 한달에 평균적으로 가입한 회원수는 82명, 168명인 반면에, 2019년도에는 68명이었다. 2019년도에 가입한 회원 수가 현저하게 줄어든 편이다. 그렇기 때문에 회원가입을 추가적으로 유도할 수 있는 해결책이 필요하다.
- 전체 고객에서 단골 고객이 차지하는 비중은 약 2%로, 해당 카페의 방문객 중 단골 고객의 비중은 낮았다. 그러나 단골 고객의 정의가 한달에 절반 이상을 방문한 고객이기 때문에, 기존 회원들의 충성도를 높이는 방안을 모색해 볼 수 있을 것 같아 분석을 진행하였다.

2.1 단골 고객들의 주 성별 및 연령대 분석



[차트32] 단골 고객의 주요 연령대

단골 손님들의 연령대 분포만 보았을 때, 20 대의 비율이 압도적으로 높은 것을 알 수 있다.



[차트 33] 단골 고객의 연령대 별 성별

그러나, '성별'이라는 변수를 추가하여 그래프를 그렸을 때에는 40 대 여성의 수가 훨씬 더 높았다. 그러나 20 대에서 성별을 'N'으로 표시한 회원들의 비율이 높기 때문에, 40 대 여성 고객의 비율이 높다고 할 수 없다. 그렇기 때문에 단골 고객의 선호 메뉴와 주로 이용하는 시간대에 대한 통계를 낼 때는 단골 고객의 연령대를 기준으로 집계하였다.

시간대에 대해 데이터를 집계 하기 전, 아래 기준으로 시간 데이터를 전처리를 진행하였다.

시간	명칭
오전 6 시~오전 11 시	아침(morning)
오전 11 시~오후 2 시	점심(lunch)
오후 3 시~오후 5 시	오후(afternoon)
오후 6 시~	밤(night)

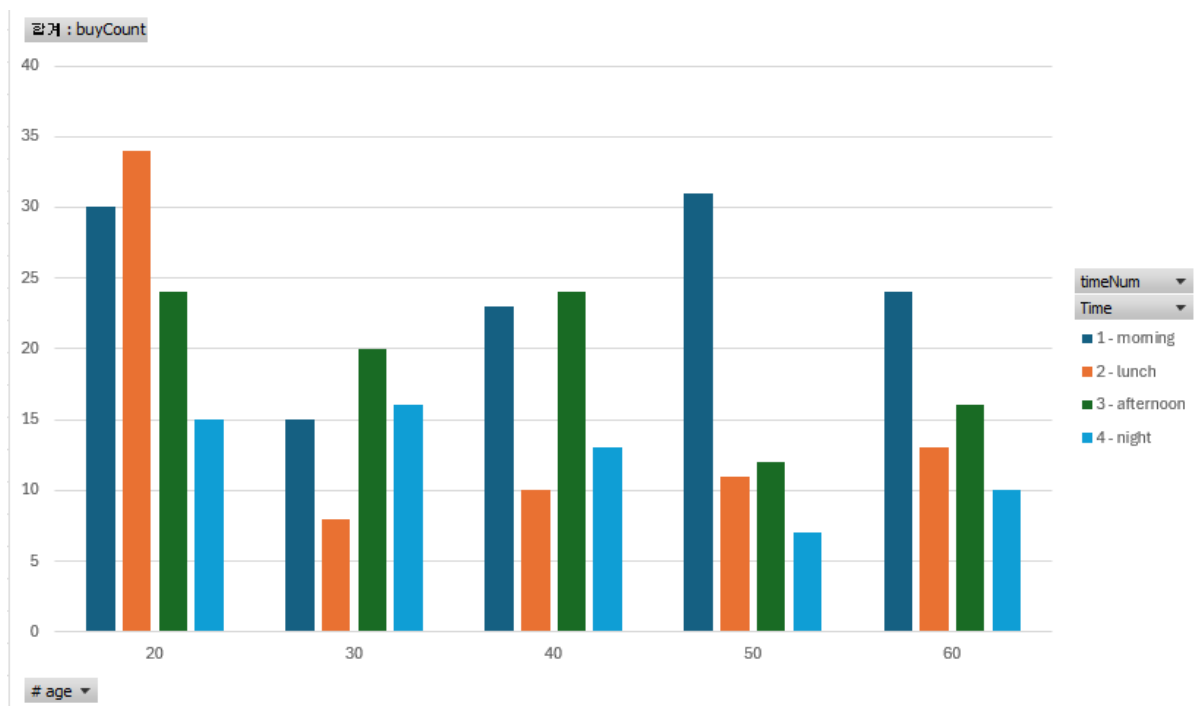
연령대의 경우 아래와 같은 기준으로 데이터를 전처리 하였다.

기준	명칭
2000 년생 이후	10 대
1990 년생 이후	20 대
1980 년생 이후	30 대

1970 년생 이후	40 대
1960 년생 이후	50 대
1950 년생 이후	60 대

[도표 11] 시간, 연령대 데이터 전처리

2.2 단골 고객들이 주로 이용하는 시간대



[차트34] 단골고객의 연령대별 구매 시간대

- 20 대 단골 고객들은 '점심' 시간대에 가장 많이 구매 하는 것을 확인할 수 있었다. 그러나 '아침'인 경우와 4 건 차이가 나기 때문에, '아침' 시간대에도 많이 구매한다고 볼 수 있다.
- 30 대, 40 대 단골 고객들은 '오후' 시간대에 가장 많이 구매 하는 것을 확인할 수 있었다.
- 그러나 5-60 대 단골 고객들의 경우 '아침' 시간대에 가장 많이 구매를 하는 것을 확인할 수 있었다.

2.3 단골 고객들이 자주 마시는 음료 분석

20 대

시간대	품목명	판매 수
아침	Barista Espresso	30
점심	Barista Espresso	17
	Brewed Chai Tea	17
오후	Barista Espresso	24
밤	Brewed Chai Tea	15

30 대

시간대	품목명	판매 수
아침	Gourmet brewed coffee	15
점심	Organic brewed coffee	8
오후	Brewed Chai tea	10
	Brewed herbal tea	10
밤	Barista Espresso	8
	Brewed Chai tea	8

40대

시간대	품목명	판매 수
아침	Barista Espresso	15
점심	Brewed Black tea	8
오후	Barista Espresso	10
	Hot chocolate	10
밤	Barista Espresso	13

50대

시간대	품목명	판매 수
아침	Gourmet brewed coffee	31
점심	Brewed Black tea	11
오후	Brewed Chai tea	12
밤	Brewed Chai tea	7

60대

시간대	품목명	판매 수
아침	Gourmet brewed coffee	31
점심	Brewed Black tea	11
오후	Brewed Chai tea	12
밤	Brewed Chai tea	7

[도표12] 연령 및 시간대별 판매가 높았던 음료

- 20대와 40대 단골 손님들은 아침에 Barista Espresso를 선호하나 30대,50대,60대는 Gourmet brewed Coffee를 선호하는 경향을 보였다.
- 30대, 50대, 60대의 경우 대체적으로 'Brewed' 라고 붙어 있는 음료, 즉 끓인 음료를 선호하는 것을 알 수 있었다.
- 20대의 경우 'Barista Espresso' 또는 'Brewed Chai Tea'를 전반적으로 선호하는 것을 확인할 수 있었다.
- 40대의 경우 'Barista Espresso'를 선호하는 편이었다.

2.4 비 단골 회원들이 선호하는 시간대 및 선호 메뉴

위에서 그렸던 boxplot 안에서 1분위~3분위 사이에 속하는 손님들이 주로 어느 요일, 어느 시간대에 음료를 구매하였는지에 대해 분석을 진행하였다. 음료를 많이 사 마셨다고 할 수는 없지만, 회원 등록을 하고 한달에 6~11회 정도 동일한 카페에서 사 마시는 거라면 특별한 이유가 있을 것이라고 생각했기 때문이다.

연령대	선호 시간대	주로 먹는 메뉴
-----	--------	----------

10대	월요일 아침	Gourmet brewed Coffee
20대	월요일 아침	Gourmet brewed Coffee, Brewed Chai tea
30대	월요일 아침	Brewed Chai Tea
40대	금요일 아침	Barista Espresso
50대	수요일 아침	Gourmet brewed coffee
60대	수요일 아침	Barista Espresso

[도표13] 비단골 회원 연령대별 별 선호 시간대 & 메뉴

- 분석 결과, 주로 젊은 연령층 (10대~30대) 들은 월요일 아침을 선호한다는 것을 알 수 있었다. 반대로 40대는 금요일 아침에, 50대와 60대는 수요일 아침을 선호한다는 것을 확인할 수 있었다.
- 10대는 Gourmet brewed coffee, 20대는 Gourmet brewed Coffee 및 Brewed Chai tea를 선호하였고, 30대는 Brewed Chai Tea를 선호하는 것을 확인할 수 있었다.
- 40대와 50대의 경우 Gourmet brewed coffee를 선호하였고, 60대의 경우 Barista Espresso를 선호하는 것을 알 수 있었다.

2.5 비회원 매출 분석

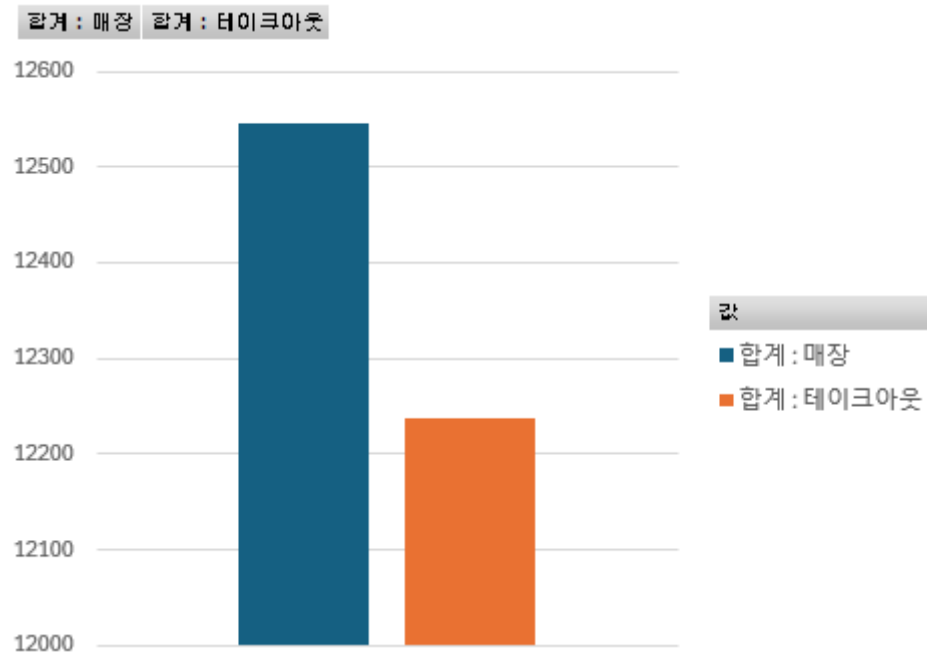
2.5.1 시간대별 선호 메뉴

시간	메뉴	주문 수
아침	Barista Espresso	2582
점심	Gourmet brewed Coffee	713
오후	Gourmet brewed Coffee	699
밤	Gourmet brewed coffee	413

[도표14] 비회원 시간대별 선호 메뉴

- 비회원들은 아침에는 'Barista Espresso'를 선호하지만, 점심/오후/밤 시간대에는 'Gourmet Brewed Coffee'를 선호한다는 것을 알 수 있었다.

2.5.2 테이크아웃 선호 여부



[차트35] 비회원 대상 매장/테이크아웃 여부

- 비회원 매출을 분석한 결과, 매장에서 먹고 가는 고객이 약 300건 정도 높았다. 이는 매장에 1회성으로 들리는 고객의 비중이 높다는 것을 알 수 있다.
- 한편, 테이크아웃 해가는 고객의 비중이 매장에서 먹고 가는 고객에 비해 현저하게 낮다고 할 수는 없다. 이 점을 바탕으로 회원 가입을 유도하는 전략에 대해 생각해 보았을 때, 테이크아웃 해 가는 비회원 고객을 대상으로 하는 전략을 세우면 효과적일 것으로 판단된다.

2.6 가설 검증 결과

- 초기에 가설을 세울 때는 2019년도에 가입을 한 회원들의 비중이 높을 것이라고 예측하였으나, 데이터를 통해 검증을 진행하니 2018년도에 가입을 한 회원들이 이용한 비중이 높았다. 그 중에서도 매출을 낸 회원들은 20대 회원들이 가장 많았다.
- 연령대 별로 선호하는 음료가 다양한 것으로 보아, 해당 카페의 특정 메뉴에 대한 선호도는 쏠리지 않은 것을 알 수 있었다.

- 단골 손님 연령대 별로 선호 방문 시간이 달랐는데, 연령대 별로 동선 및 방문 목적이 매우 다양할 것으로 추측이 된다.
- 그러나 비 단골 손님의 경우에는 '아침' 시간대를 선호하였다. 또한 젊은 연령대는 월요일을 선호한 반면, 중장년층 연령대는 평일인 수요일, 금요일을 선호하였다.
- 또한 20대, 40대 단골 손님의 경우에는 'Barista Espresso'를 선호하는 경향을 보였으나, 비 단골 손님들은 대체적으로 'Brewed' 메뉴를 선호하는 것을 확인할 수 있었다.
- 대체적으로 비 단골 손님들은 'Brewed' 메뉴를 선호하였으나, 비회원 손님들은 아침에는 'Barista Espresso', 나머지 시간대에는 'Gourmet brewed Coffee'를 선호한다는 것을 알 수 있었다.

프로젝트 결론

1. 지점 및 매출 목표 달성도 분석 결론 및 전략 제안

- 총 표본의 크기는 현지 거주 인구보다 유동인구가 배로 많다. 타 지역보다 주거 인구가 많은 아울렛(3)의 경우 프로모션 없이도 총 거래 목표를 달성했다. 그에 반해 유동인구가 많은 아울렛(5)와 아울렛(8)은 프로모션의 결과로 총 거래 목표를 겨우 달성하거나 달성하지 못하였다. 이를 통해 해당 브랜드가 관광객 보다 는 현지인들에게 접근성이 비교적 좋음을 유추해볼 수 있었다. 이에 우리나라 브랜드의 경우 관광객 등 유동 인구를 대상으로 한 프로모션 진행 및 마케팅 전략수립을 펼치는 방안을 고려해보면 좋다.
- 현지브랜드 매출 구성은 매장 취식 51%, 포장 49%로 유동인구를 고려한 마케팅 방안으로 포장시 일부 금액 할인 등의 프로모션을 제안한다. 이를 통해 매장 실내 혼잡도가 줄어들어 쾌적한 환경을 제공하고, 유동 인구를 대상으로한 매출 증가를 기대해볼 수 있다.
- 카테고리별 달성도를 고려하면 원두, 식품 등은 타 상품과의 연계를 통해 강화해야되는 부분이며, 특히 식품 카테고리 내 초콜렛칩이 들어간 상품을 다각화 하여 운영하면 매출이 증가할 수 있다.
- MD 상품들은 초기 진입시 재고를 ¼ 수준으로 운영하거나, 선오더후생성 방식, 항시 운영이 아닌 시즌성 상품 운영 등을 고려하여 재고 최적화 운영이 필요하다

다.

2. 주요 상품 데이터 분석 결론 및 전략 제안

- 시간대별 주요 상품 분석의 경우, 모든 상품 카테고리에서 오전 판매량 대비 오후 판매량이 급격히 감소하는 경향이 나타났다. 이는 주로 아침 식사와 함께 커피를 마시는 미국 커피 문화에서 기인된 것으로 보인다. 그럼에도 불구하고 여전히 오후를 포함한 모든 시간대에서 커피가 다른 상품보다 높은 매출을 기록하고 있었다. 따라서 고객들이 오후에도 부담 없이 커피를 즐길 수 있도록 디카페인 커피를 출시하는 전략을 제안한다.
- 티는 커피 다음으로 인기가 많은 상품 카테고리였다는 점을 고려할 때, 티의 판매량과 매출을 커피 수준으로 끌어올리기 위한 전략이 필요하다. 이를 위해 티를 좋아하는 고객층을 대상으로 맞춤형 혜택을 제공하여 티 카테고리를 커피와 동등한 주력 상품으로 성장시켜야 한다.
- 'Chocolate Croissant'은 판매량과 매출 모두에서 가장 높은 수치를 기록하고 있으므로, 디저트 상품으로 가장 주력해서 판매해야 할 상품이다. 반면, 'Oatmeal Scone'은 판매량과 매출 모두에서 낮은 수치를 기록하고 있고, 'Ginger Scone'은 프로모션을 했음에도 버려지는 비율이 70% 이상이므로, 국내 브랜드 진출 당시에는 출시하지 않는 전략을 제안한다.
- 베이커리는 유통기한이 짧은 특성으로 인해 당일 폐기를 해야 하므로, 수요일에 비해 과도한 재고가 충당될 시 많은 쓰레기가 발생할 뿐 아니라 손해가 막대할 수 있다. 따라서 국내 브랜드 진출 초기에 수요를 정확하게 파악하여 재고를 점차 조율해 나갈 것을 제안한다.
- 베이커리와 함께 가장 자주 구매된 음료의 특징은 '부드러운 커피'였다. 국내 브랜드 진출 초기에 판매량을 증대시키기 위한 방안으로, 베이커리와 부드러운 커피와의 조합을 중심으로 한 번들 프로모션을 진행하는 전략을 제안한다.
- 건강관리에 관심이 높아진 트렌드에 따라 'Sugar Free Syrup'이 꾸준히 많은 판매량을 보였다. 이에 따라 새로운 저당 음료를 출시할 수 있을 것이고, 고객들의 주목을 더욱 끌어들이기 위해 메뉴판에 'Low Sugar' 또는 'Zero Sugar'와 같이 별도 표기하는 전략을 제안한다.
- 'Latte'에 'Caramel syrup', 'Hazelnut syrup', 'Chocolate syrup'을 넣어 먹는 조합이 많은 것으로 나타났다. 해당 조합에 대한 고객들의 높은 선호도에 따라

시럽을 첨가해서 먹는 것이 아닌 'Caramel Latte', 'Hazelnut Latte', 'Chocolate Latte'와 같이 하나의 메뉴로 출시하는 전략을 제안한다.

- 개별적으로 프로모션 되었던 'Ginger Scone'과 'Ouro Brasileiro shot'을 함께 구매하는 경향이 높았던 점을 고려할 때, 번들 프로모션을 도입하는 방식도 고려해볼 수 있다. 따라서, 개별 프로모션보다 더 큰 시너지를 창출할 수 있는 번들 프로모션을 통해 전체 매출을 극대화하는 전략을 제안한다. 특히 가장 인기가 높았던 베이커리인 'Chocolate Croissant'와, 베이커리와 조합에서 선호도가 높았던 'Colombian Medium Roast'를 번들로 묶는 프로모션을 추진하는 것이 효과적일 것이다.

3. 재방문 고객 데이터 분석 결론 및 전략 제안

- 각 매장에는 한 달에 2회 이상 방문하는 고객의 비율이 매우 높았고, 그 중 6회 이상 방문한 고객의 수는 약 70%로 각 매장별로 반복적으로 방문하는 단골 고객이 있음을 확인할 수 있다.
- 한 고객이 여러 매장을 다닐 수 있는 상황을 가정해 customer 데이터를 조회한 결과, 서로 다른 customer_id에 같은 사람의 정보가 저장되어 있는 경우는 없음을 확인했다. 따라서 한 customer는 한 outlet만 이용했다는 것을 알 수 있다.
- 각 매장별로 고객들의 방문횟수 정보와 누적 매출로 파레토 곡선을 그린 결과, 상위 20%의 고객들이 전체 매출의 40%씩 가까이 차지하고 있음을 알 수 있다.
- 따라서 단골 고객들의 주요 구매 품목을 파악하는 것이 매출에 영향을 끼칠 것이다.
- 각 매장별로 단골 고객들이 많이 구매한 메인 메뉴 품목을 4개까지 집계한 결과, outlet3에서는 'Earl Grey Rg', 'Dark chocolate Lg', 'Peppermint Rg', 'Sustainably Grown Organic Lg'의 판매량이 많았고, outlet5에서는 'Jamaican Coffee River Lg', 'Ethiopia Rg', 'Latte Rg', 'Cappuccino Lg'의 판매량이 많았고, outlet8에서는 'Ouro Brasileiro shot', 'Ethiopia Lg', 'English Breakfast Lg', 'Brazilian Lg'의 판매량이 많았다.
- 따라서 위와 같은 상품을 중심으로 메뉴를 구성하고 프로모션을 진행한다면, 국내 브랜드의 해외 진출 시 매출을 상승 및 유지에 큰 기여를 할 것으로 기대된다.

4. 회원 데이터 분석 결론 및 전략 제안

- 해당 카페는 회원 비중의 매출이 높은 곳이다. 회원의 가입 년도는 2017년도 ~2019년도로 범위가 구성되어 있으며, 매출의 대부분은 2018년도 회원을 가입한 고객들의 구매 데이터로 구성되어 있었다.
- 년도별 회원가입자 수를 분석해 보았을 때, 2017년도 및 2018년도 한달에 평균적으로 가입한 회원수가 82명, 168명인 반면에, 2019년도에는 68명이었다. 즉, 2019년도에 가입한 회원 수가 현저하게 줄어든 편이다. 그렇기 때문에 회원가입을 추가적으로 유도할 수 있는 해결책이 필요하다.
- 회원 혜택 부여 여부는 매출 데이터에서 확인하기는 어려웠다. 이 점을 보았을 때, 회원 가입 혜택이 크지 않았다는 것을 알 수 있었다.
- 회원들은 평균 6~11일 정도의 방문 횟수를 지니고 있었다.
- 데이터 분포도를 그렸을 때, 18회 이상 방문한 고객들이 존재하였다. 해당 고객들은 전체 회원수의 2%에 속한다. 그러나 18회 이상 방문한 고객들은 분포를 벗어난 통계였기 때문에, 해당 보고서에서 '단골'로 정의하였다.
- 단골 손님은 대부분 20대 였는데, 아침~점심 시간대에 자주 방문하는 것을 알 수 있었다.
- 비 단골 손님들을 대상으로 분석을 해보면, 20~30대의 젊은 연령층들은 월요일 아침에 방문을 하는 패턴을 보였으며, 4~60대의 연령층들은 수, 금요일 아침에 방문하는 패턴을 보였다.
- 단골 손님들은 연령대별로 선호하는 메뉴가 달랐으나, 비 단골 손님들은 'brewed'가 붙은, 즉 커피 원두나 차 등을 우리는 메뉴를 선호하는 경향이 강했다.
- 이러한 결론을 바탕으로 국내 브랜드가 진출하였을 때 회원 유치 방향을 생각해 보자면, 20대 연령층을 대상으로 회원 혜택을 부여하는 방향을 확인할 수 있다.
- 또한 월요일 아침에 회원에게 할인을 주는 방향을 생각해 볼 수 있다. 하지만, 다른 연령대 손님들은 월요일 외 다른 요일을 선호하는 경향이 있다. 그렇기 때문에 다른 날 아침에도 회원에게 'brewed' 메뉴를 대상으로 할인을 할 수 있게 하는 방향이 있을 것이다.
- 또한 20대 단골 손님들은 아침~점심 까지 자주 구매하는 패턴을 보이고 있는데, 단골 회원에게는 디저트를 증정해 주는 방향도 고려해 볼 수 있을 것이다.

- 이는 '주요 상품 데이터 분석' 에서 현재 이슈로 삼고 있는 'Bakery 재고 이슈'를 해결하는 효과도 낼 수 있을 것으로 기대한다.
- 비회원 데이터를 분석해 보았을 때, 매장 내에서 음료를 마시고 가는 경우가 테이크아웃에 비해 높은 편이었다. 그러나 이는 약 500건 정도 차이 나는 수치이므로, 비회원 중에는 테이크아웃을 하는 손님들도 많았다는 것을 알 수 있다. 또한 아침에 'Barista Espresso' 를 마시는 것을 선호하는 것도 확인할 수 있었다. 또한 나머지 시간대에는 'Gourmet brewed Coffee' 마시는 것을 선호한다는 것을 알 수 있었다.
- 그렇기 때문에 회원 가입을 하면 테이크아웃 Barista Espresso 메뉴를 아침마다 할인을 해주는 혜택을 부여하는 방법을 제안한다.
- 또한 나머지 시간대에는 스타벅스의 '오늘의 커피' 메뉴 처럼 매 요일 마다 원두를 다르게 해서 'Gourmet brewed Coffee' 메뉴를 할인하여 회원들에게 제공하는 혜택을 부여 하는 방법도 좋을 것이라고 생각이 든다.

한계점

- **2019년 4월 1일~2019년 4월 6일까지 데이터 누락 가능성**
 - 2019년 4월 1일~2019년 4월 6일 까지 회원 매출만 존재하고 있다. 그러나 카페라는 장소 특성상 비회원도 들릴 수 있는 장소이기 때문에, 비회원 매출이 누락이 되어있을 가능성이 있다.
- **데이터 갯수의 한계**
 - 이번 프로젝트에 사용된 'Coffee shop sample data' 데이터는 2019년 4월 한달 만의 데이터를 포함하고 있어서 데이터의 분석 범위가 제한되고, 고객의 충성도나 반복 구매 패턴을 명확하게 파악하는 데에 한계가 있다.
 - 또한 고객의 장기적인 구매 패턴이나 매출 변동을 파악하기가 힘들다.
- **매출에 영향을 줄 수 있는 타변수의 부재**
 - 프로모션 여부를 알 수 있는 컬럼은 존재했지만, 어떤 프로모션을 진행했는지에 대한 상세정보를 알 수 있는 테이블의 부재로 프로모션이 매출에 끼치는 영향을 파악하기에 제한이 있다.
 - 매출은 단순히 카페 내부의 메뉴나 프로모션만이 아닌 날씨나 특정 기간에 열리는 행사와 같은 외부 변수에도 영향을 받는데, 해당 데이터를 사용하지 못해 매출에 대한 명확한 인사이트를 제공하는데에 한계가 있다.

- 사용한 데이터가 익명화된 카페데이터여서 경쟁업체가 될 수 있는 타 매장과
의 비교가 힘들었기 때문에, 주어진 데이터 셋 내에서만 분석이 가능했다.