

Tableau DataFest



모든 사용자를 위한 고급 분석 (AI / ML)

테블로 잘 알려주는 예쁜 누나

박소영

Solution Engineer

Tableau

#Data22



사용자의 페르소나에 따라 활용 가능한 분석 영역이 달라집니다.

비즈니스 사이언티스트
가이드드 & 자동화 ML

No-code ML Model Authoring

Simulation



Scoring in Data Pipelines



Prediction Monitoring

Visual Advanced Analytics &
Predictive Calculations



BI 분석가 & 익스플로러
AI-증강 분석

Discovery Stories
Explain Data-
Report Insights



Ask Data
Conversational Queries



End Users
Better, Faster
Decision-making with AI

데이터 사이언티스트
사용자정의 모델 구축 및 배포



BYOM



Analytics Extensions



Notebooking



Prep Script Nodes

Tableau Native Features

Explain Data, Clustering



사용자의 페르소나에 따라 활용 가능한 분석 영역이 달라집니다.

비즈니스 사이언티스트
가이드드 & 자동화 ML

No-code ML Model Authoring

Simulation



Scoring in Data Pipelines



Prediction Monitoring

Visual Advanced Analytics &
Predictive Calculations



BI 분석가 & 익스플로러
AI-증강 분석

Discovery Stories
Explain Data-
Report Insights



Ask Data
Conversational Queries



End Users
Better, Faster
Decision-making with AI

데이터 사이언티스트
사용자정의 모델 구축 및 배포



BYOM



Analytics Extensions



Notebooking



Prep Script Nodes

데이터 설명 Explain Data

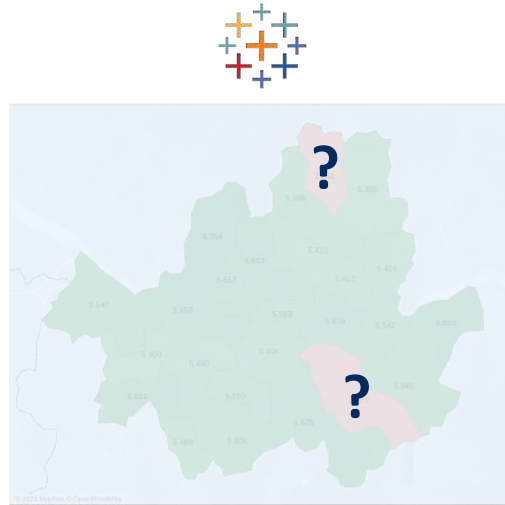


“데이터 설명”은 무엇인가요?

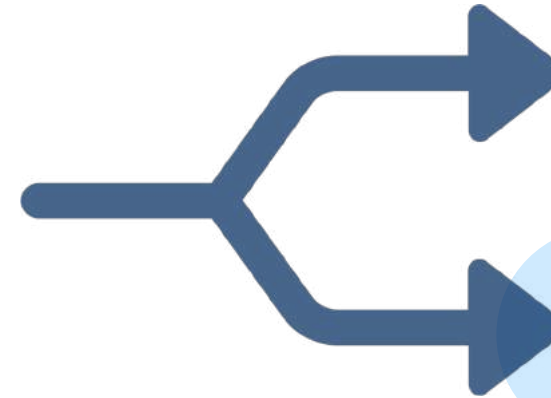
다음으로 탐색해봐야 할 경로를 추천하는 “네비게이션”



지역별 고객만족도는?



특정 지역의 만족도가 낮은 이유가 무엇일까?



조사 인원 수 자체가 적은 것은 아닐까?

해당 지역 담당자의 연차가 낮은 것은 아닐까?



데모 시나리오

주류 회사 영업 사원의 고객 만족도 평가 분석

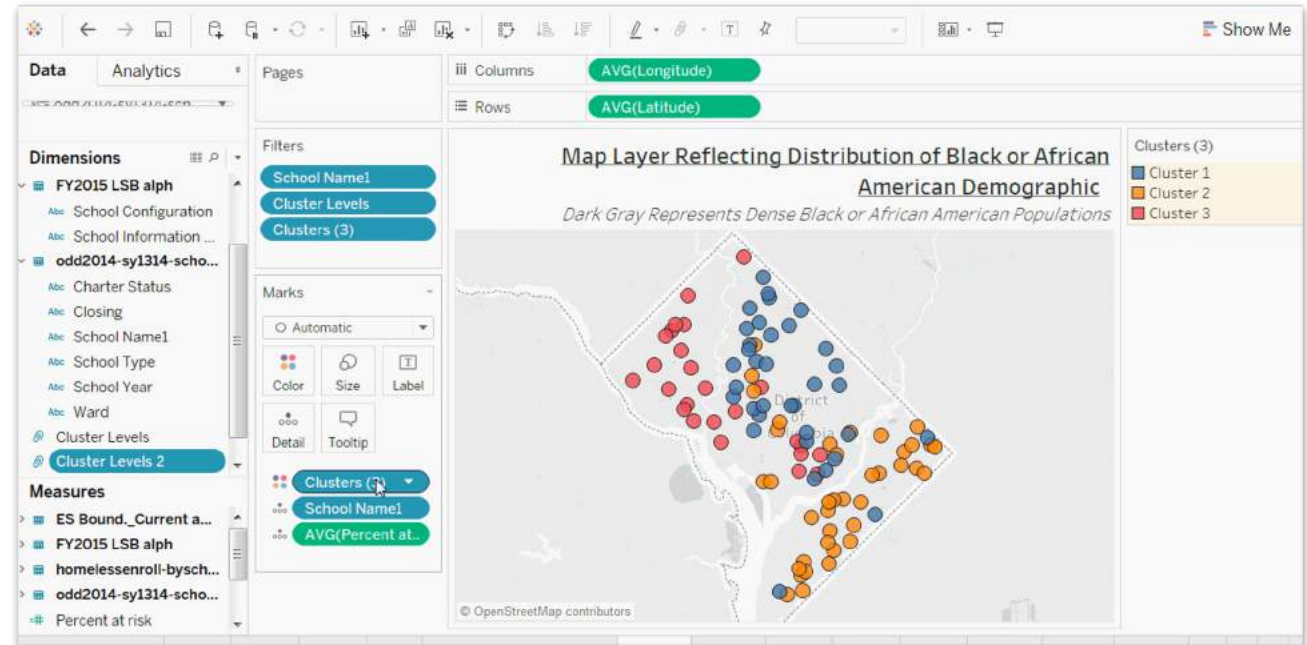
고객id	영업사원id	영업연차	시도	시군구	업종	Index	고객만족도	연속영업기간 (개월)
24	10	< 5	서울특별시	금천구	음식점	2	8	21
2	3	< 5	서울특별시	성북구	편의점	5	7	10
21	2	< 5	서울특별시	마포구	일반주점	11	7	23
14	3	< 5	서울특별시	은평구	편의점	12	2	23
10	3	10+	서울특별시	노원구	음식점	13	6	32
22	9	5 to 10	서울특별시	영등포구	일반주점	14	10	15
17	9	10+	서울특별시	중구	유흥업소	17	9	41
11	8	< 5	서울특별시	서대문구	대형마트	18	7	19
1	10	5 to 10	서울특별시	노원구	일반주점	19	10	48
9	5	5 to 10	서울특별시	용산구	음식점	20	3	32
3	9	10+	서울특별시	관악구	일반주점	21	9	32
17	2	10+	서울특별시	성동구	일반주점	22	7	42
1	1	< 5	서울특별시	중랑구	음식점	27	4	6
3	10	10+	서울특별시	종로구	편의점	28	3	49
7	5	5 to 10	서울특별시	동작구	유흥업소	31	4	6
25	7	5 to 10	서울특별시	관악구	유흥업소	38	6	6
13	10	< 5	서울특별시	양천구	음식점	40	9	24
3	3	< 5	서울특별시	서대문구	편의점	42	5	45
11	5	10+	서울특별시	강서구	일반주점	43	2	43

군집 분석 Cluster Analysis

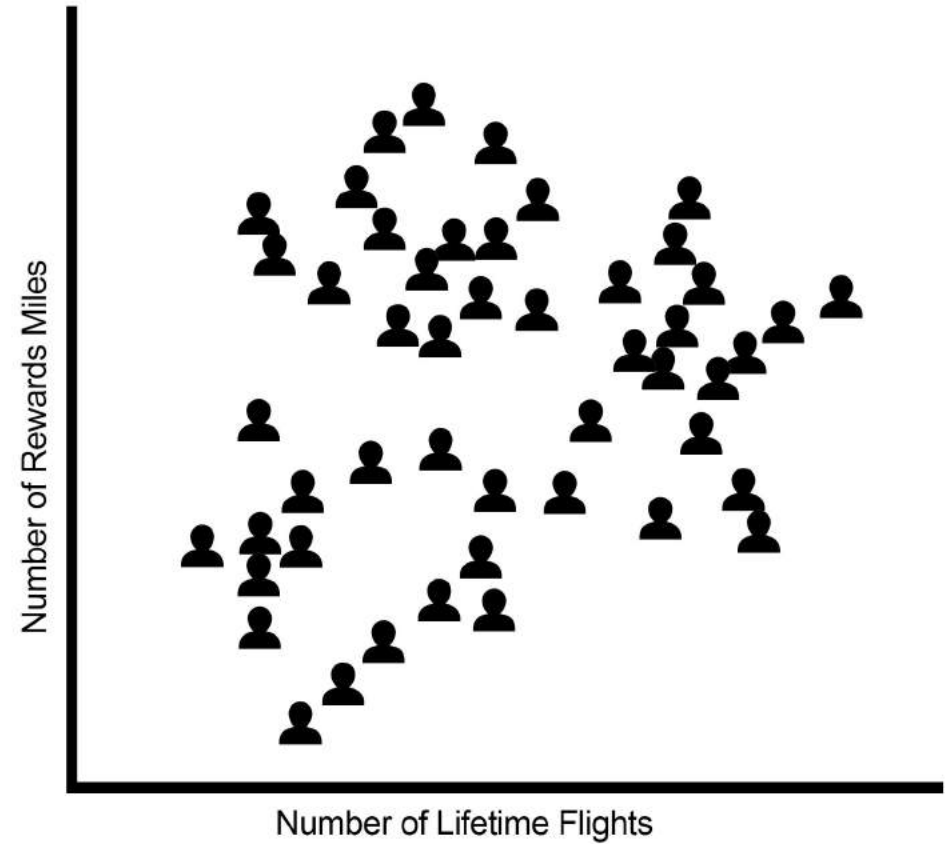


‘군집 분석’은 무엇인가요?

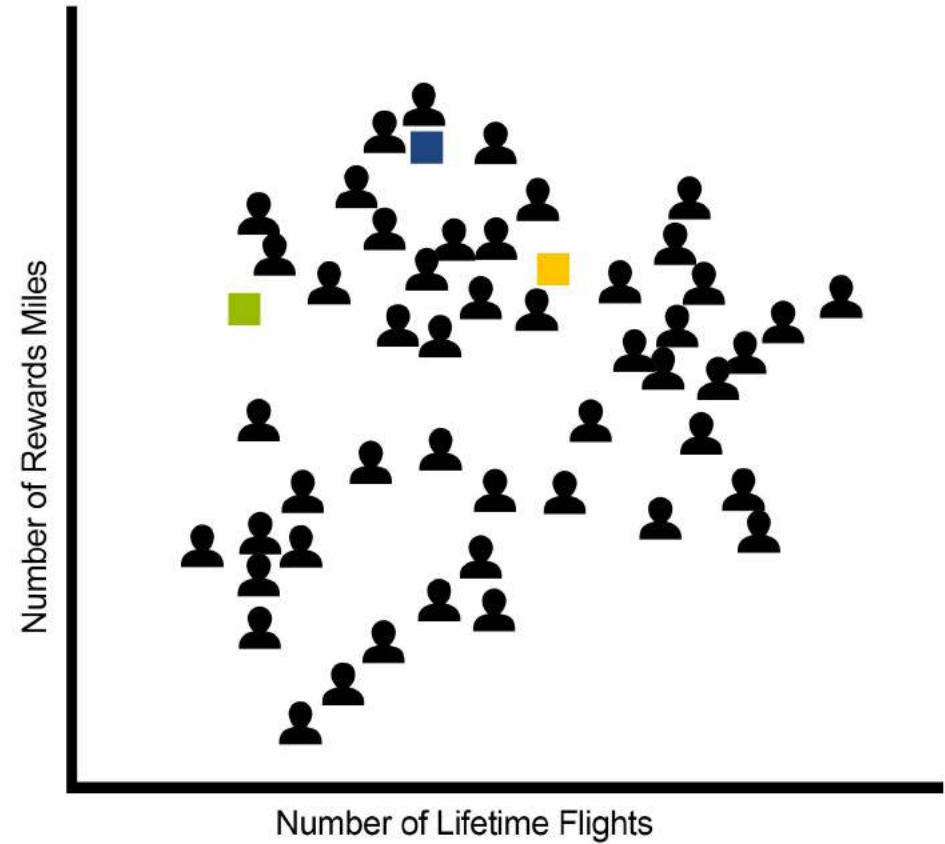
데이터에서 그룹 또는 유형을 발견해 자동으로
그룹화합니다.



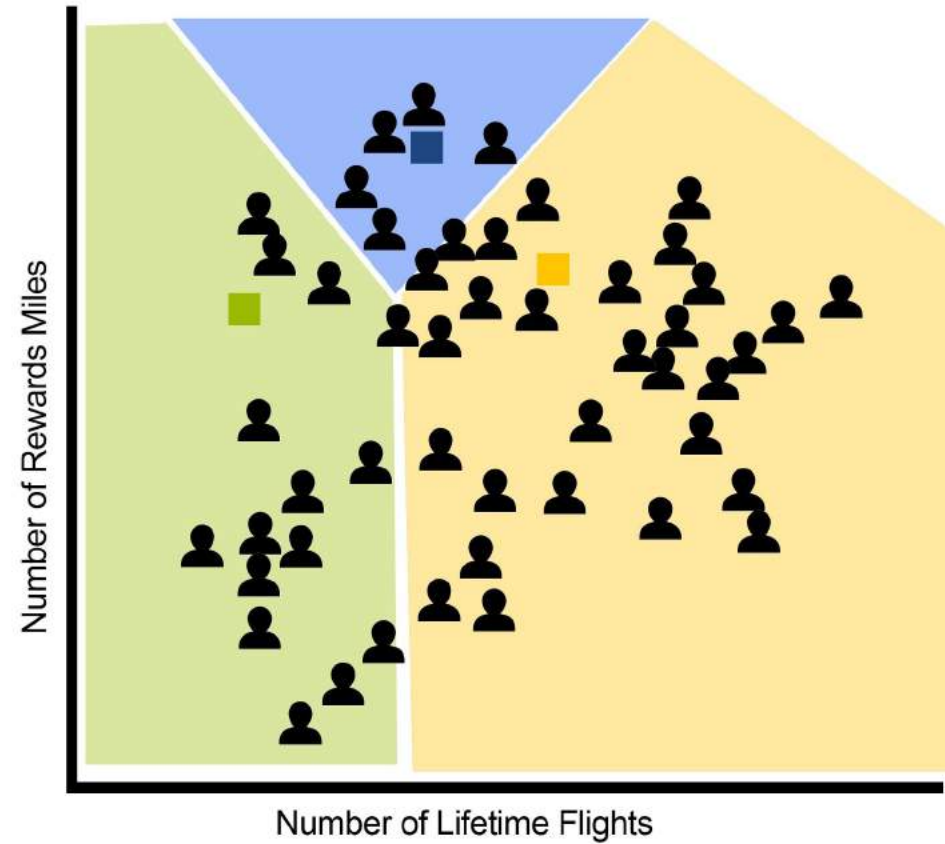
태블로가 군집을 형성 하는 방법



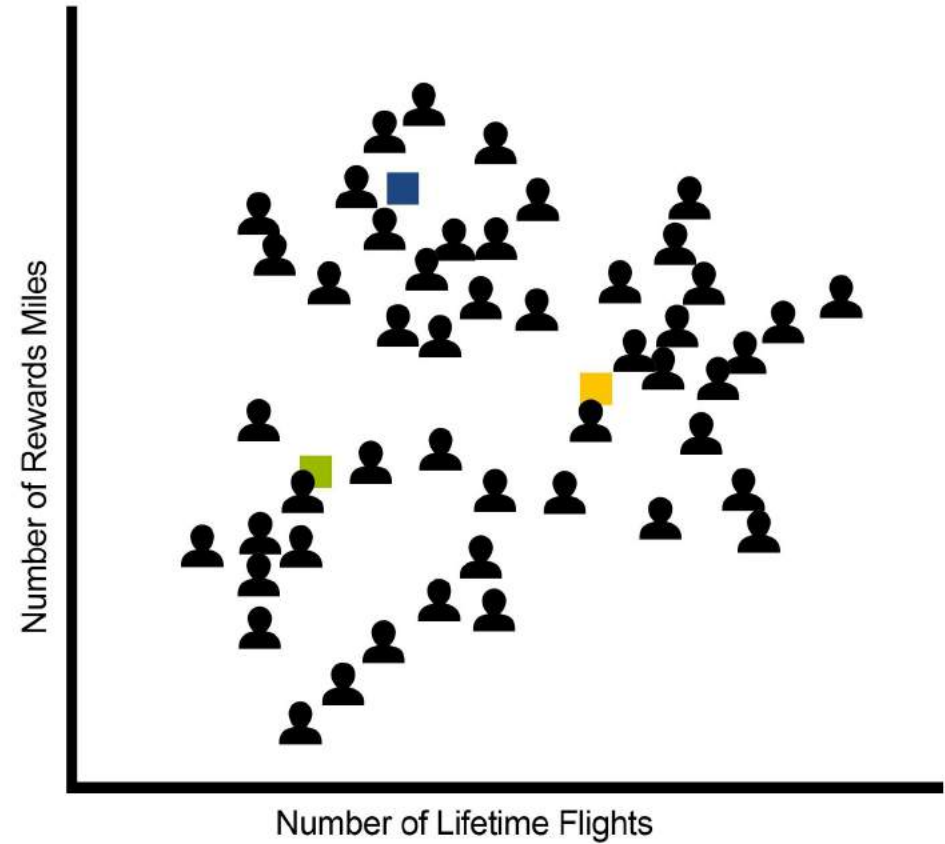
태블로가 군집을 형성 하는 방법



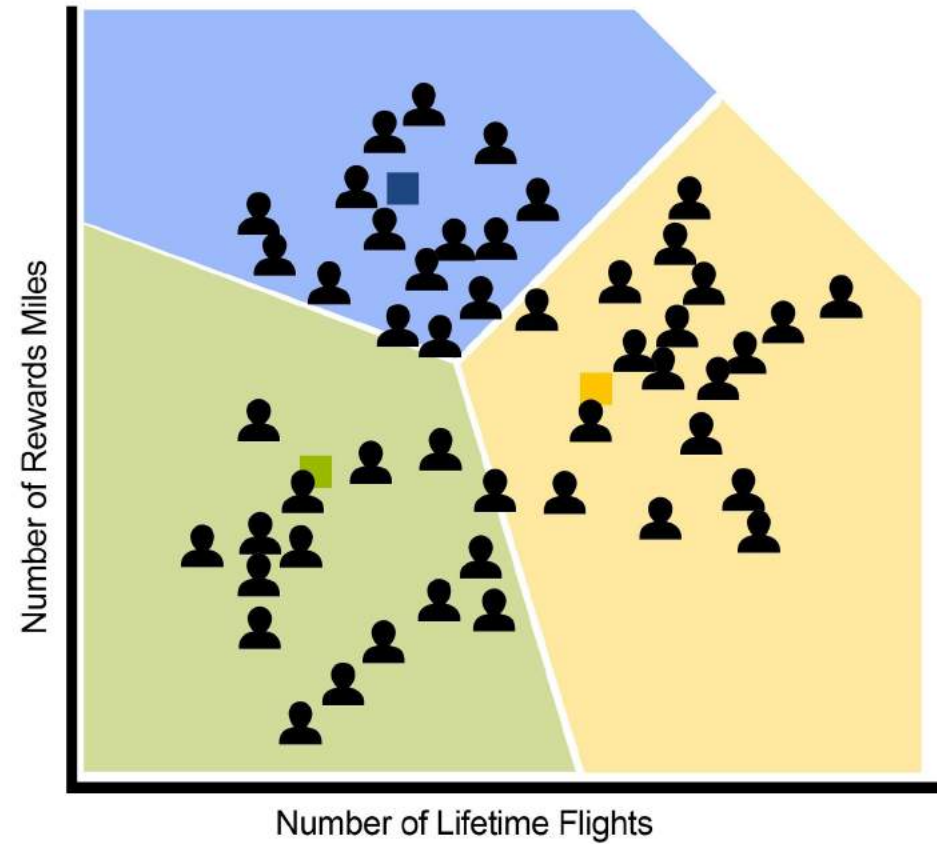
테블로가 군집을 형성 하는 방법



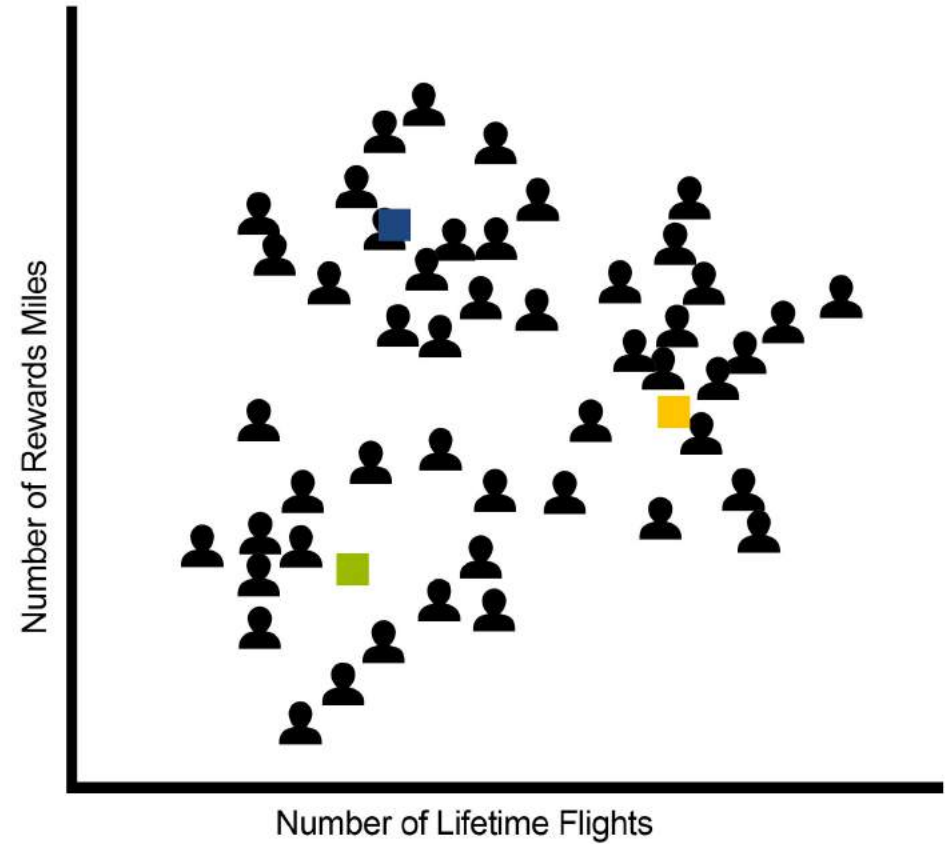
태블로가 군집을 형성 하는 방법



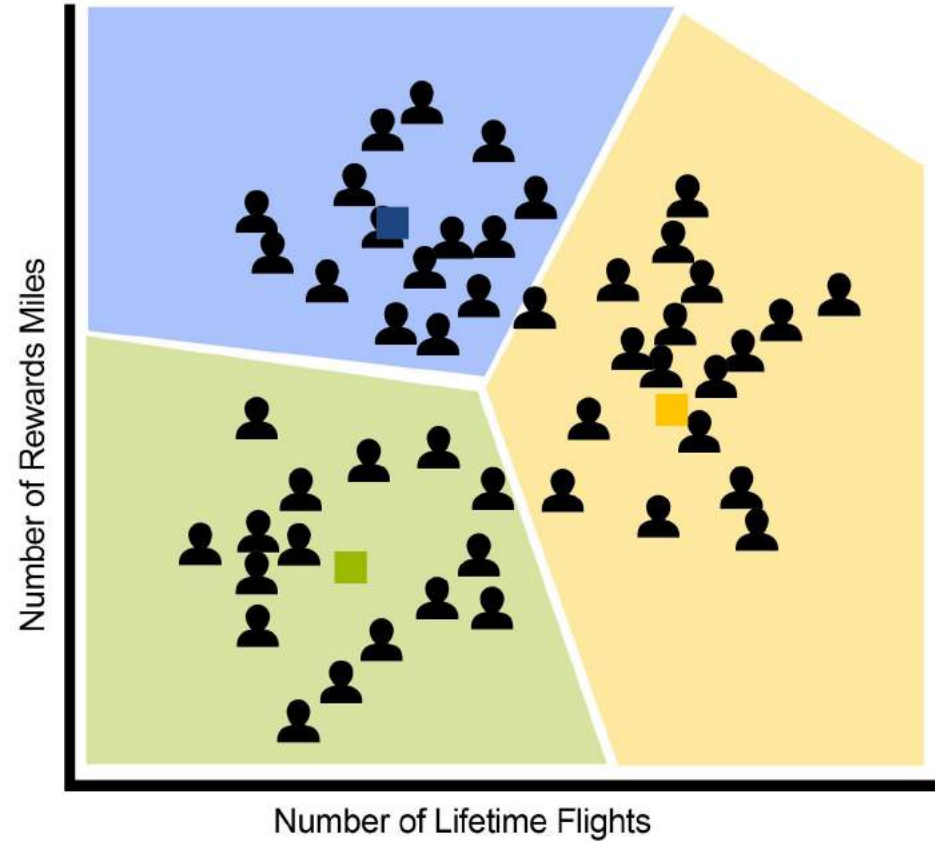
테블로가 군집을 형성 하는 방법



태블로가 군집을 형성 하는 방법



테블로가 군집을 형성 하는 방법



데모 시나리오

서비스업종별 특성에 따른 클러스터링

기준 년 코드	기준 분기 코드	상권 코드	상권명	서비스 업종 코드	서비스업종명	연령대	건수	시장성 (추정 매출액)
2021	4	2130169	대조시장	CS300033	철물점	20대	56	409953
2021	4	2130174	연희사리가	CS300033	철물점	20대	354	4275821
2021	4	2130180	포방터시장	CS300033	철물점	20대	22	366154
2021	4	2130185	망원월드컵시장	CS300033	철물점	20대	48	379687
2021	4	2130195	도화동 상점가	CS300033	철물점	20대	256	2369432
2021	4	2130197	마포시장	CS300033	철물점	20대	69	2142688
2021	4	2130201	월정로시장	CS300033	철물점	20대	80	1188738
2021	4	2130208	목동캐비시장(목3동시장)	CS300033	철물점	20대	15	466851
2021	4	2130209	목사령시장(목4동시장)	CS300033	철물점	20대	59	799797
2021	4	2130210	목2동시장	CS300033	철물점	20대	486	4959795
2021	4	2130211	신정2동 골목시장(오목교중앙시장)	CS300033	철물점	20대	59	195752
2021	4	2130218	신월중앙시장	CS300033	철물점	20대	135	2012947
2021	4	2130219	까치산시장	CS300033	철물점	20대	72	707621
2021	4	2130221	남부골목시장(남부화곡시장, 남부시장)	CS300033	철물점	20대	0	0
2021	4	2130223	고척근린시장	CS300033	철물점	20대	6	47465
2021	4	2130233	남문시장	CS300033	철물점	20대	36	3645960
2021	4	2130244	대림중앙시장	CS300033	철물점	20대	47	1184720
2021	4	2130246	영등포청과시장(조광시장)	CS300033	철물점	20대	34	429542
2021	4	2130257	영등포시장기계공구상가	CS300033	철물점	20대	949	14029697
2021	4	2130269	강남시장(동작구 강남시장)	CS300033	철물점	20대	71	579535
2021	4	2130271	남성역골목시장	CS300033	철물점	20대	71	587896

기준 년 코드	기준 분기 코드	상권 코드	연령대	상주연구수
2021	4	2110001	10대	904
2021	4	2110002	10대	1692
2021	4	2110003	10대	1833
2021	4	2110004	10대	2170
2021	4	2110005	10대	1363
2021	4	2110006	10대	496
2021	4	2110007	10대	2528
2021	4	2110008	10대	1907
2021	4	2110009	10대	728
2021	4	2110010	10대	1681
2021	4	2110011	10대	399
2021	4	2110012	10대	1330
2021	4	2110013	10대	698
2021	4	2110014	10대	753
2021	4	2110015	10대	626
2021	4	2110016	10대	987
2021	4	2110017	10대	525
2021	4	2110018	10대	691
2021	4	2110019	10대	104
2021	4	2110020	10대	1662

기준 년 코드	기준 분기 코드	서비스 업종 코드	상권 코드	점포수
2021	4	CS100001	2110126	1
2021	4	CS100001	2110208	1
2021	4	CS100001	2110446	1
2021	4	CS100001	2110795	1
2021	4	CS100001	2110974	1
2021	4	CS100001	2110017	2
2021	4	CS100001	2110683	2
2021	4	CS100001	2110956	2
2021	4	CS100001	2110959	2
2021	4	CS100001	2110026	3
2021	4	CS100001	2110070	3
2021	4	CS100001	2110097	3
2021	4	CS100001	2110100	3
2021	4	CS100001	2110103	3
2021	4	CS100001	2110143	3
2021	4	CS100001	2110145	3
2021	4	CS100001	2110149	3

Tableau

파일

데이터

워크시트

대시보드

스토리

분석

맵

서식

서버

창

도움말

Tableau - 클러스터(Start)

표현 방식

데이터

분석

페이지

서울시_업종별_추정매출

검색

폴더

기타

상권명

서비스업종명

연령대

측정값 이름

건수

기하 도형

상권 면적

상주인구수

점포당 매출액

점포수

클러스터당 평균 점포수

경도(생성됨)

위도(생성됨)

측정값

매개 변수

TOP 10 기준

서비스업종명

필터

마크

자동

색상

크기

텍스트

세부 정보

도구 설명

시트 1

여기에 필드 놓기

여기에 필드 놓기

여기에 필드 놓기

데이터 원본

서울시 상권 및 업종 선택 시뮬레이션

시트 1

Soyoung Park

이 장사, 성공할 수 있을까요?

상권 및 업종 선택
시뮬레이션

조사 기준: 2021년 4분기
출처: 서울시 공공 데이터

서비스업종명
편의점

- 유사 업종
- 클러스터 1
 - 클러스터 2
 - 클러스터 3
 - 클러스터 4
 - 클러스터 5

선택한 편의점 업종은

클러스터.. 업종과 유사한 성격을 갖고 있습니다.

시장성

추정 시장성 규모
Q4, 2021

1,271.5B

유사 업종 대비
↑18.4%



기대매출

점포당 매출액
Q4, 2021

92.3M

유사 업종 대비
↑168.8%

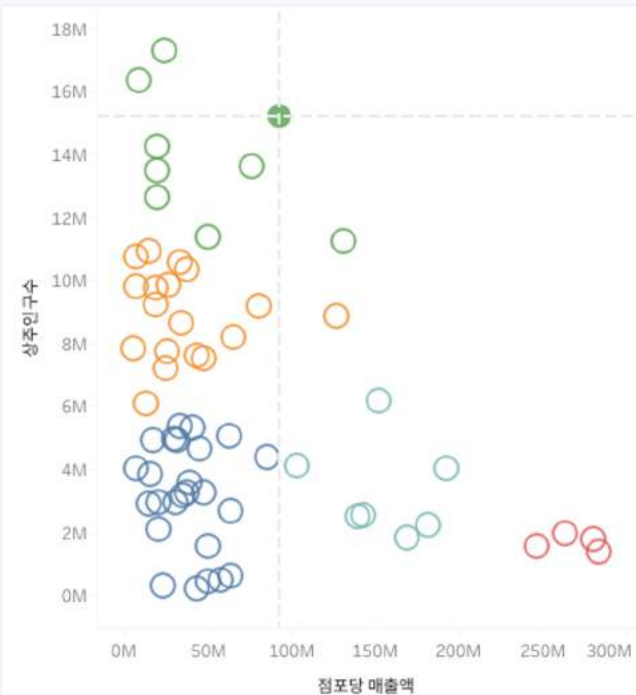


경쟁의 정도

점포수
Q4, 2021

13,770

유사 업종 대비
↓54.4%



Top 10 상권 숨기기

TOP 10 기준
시장성 (추정 매출액)

번호	상권 코드 명	5,650.5M	188.4M	30	36,800	285,287m
1	장안1동주민센터	4,292.8M	110.1M	39	20,380	331,033m
2	성수초등학교	4,203.7M	127.4M	33	34,424	204,853m
3	까치산역 3번	3,451.2M	127.8M	27	19,093	192,093m
4	논현초등학교	3,237.3M	215.8M	15	30,217	172,691m
5	한울중학교	3,181.2M	176.7M	18	32,170	72,474m
6	모래내우체국	3,180.8M	265.1M	12	23,635	305,648m
7	문래예술촌(문래창작촌)	3,156.2M	150.3M	21	19,024	236,425m
8	대정초등학교	3,133.7M	174.1M	18	10,177	181,137m
9	경리단길남측	3,127.7M	115.8M	27	31,127	257,234m
10	사당역 4번					

OG 5G 0M 200M 0 20 40 0K 20K 0K 200K

시장성 (추.. 점포당 매.. 점포수 상주인구수 상권 면적

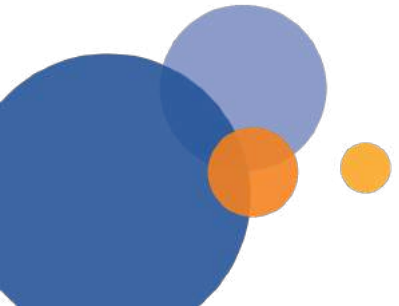
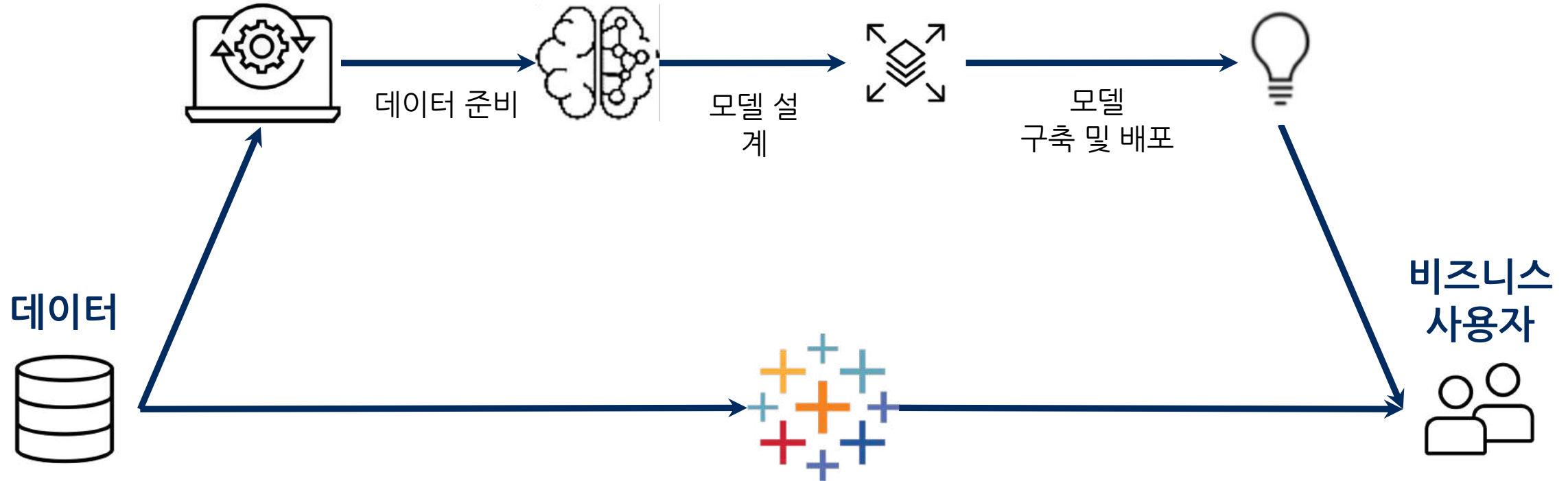
Tableau Business Science

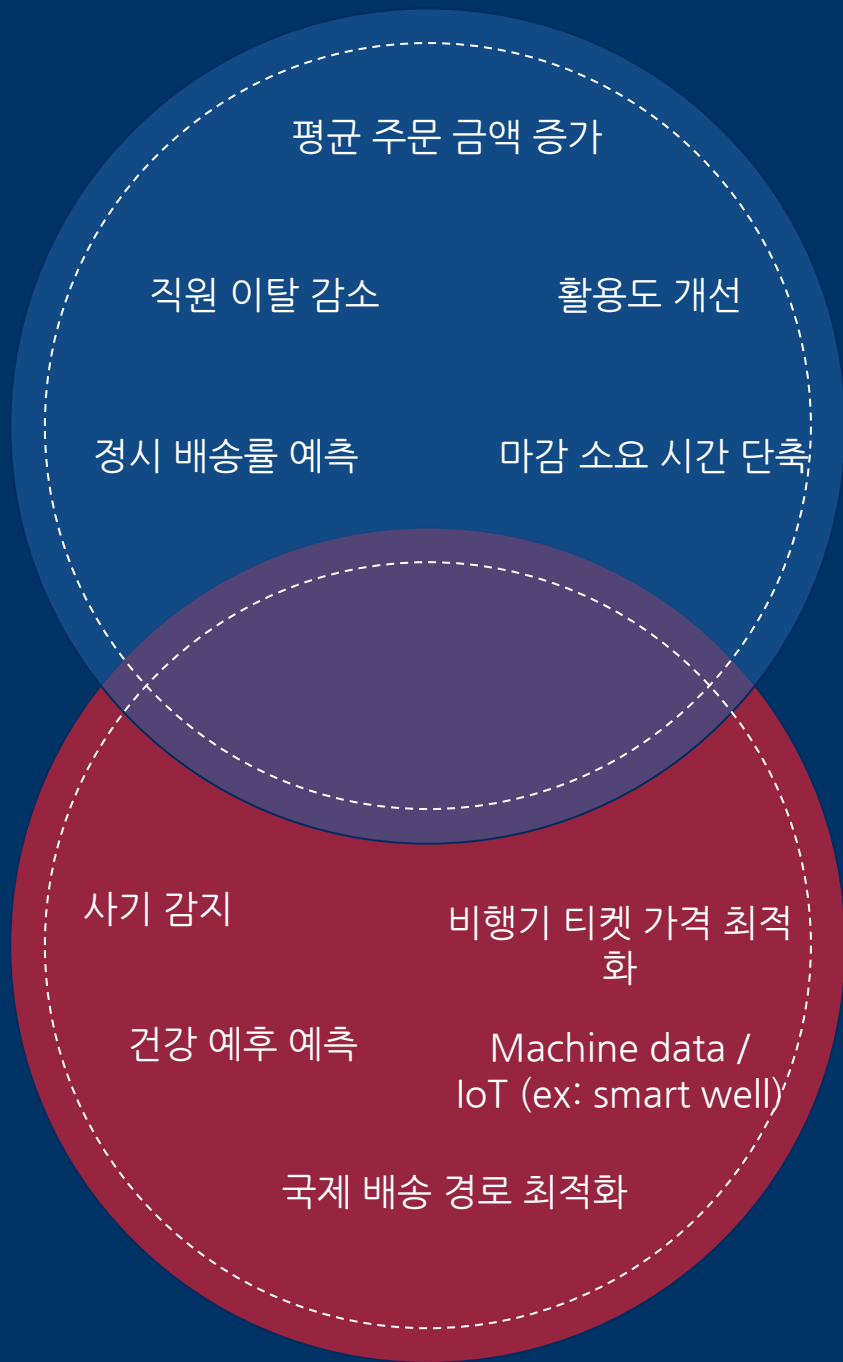


사용자의 페르소나에 따라 활용 가능한 분석 영역이 달라집니다.



Machine Learning이 Business Problem에 적용되기까지





신속함 및
컨텍스트 우선

비즈니스 사이언스



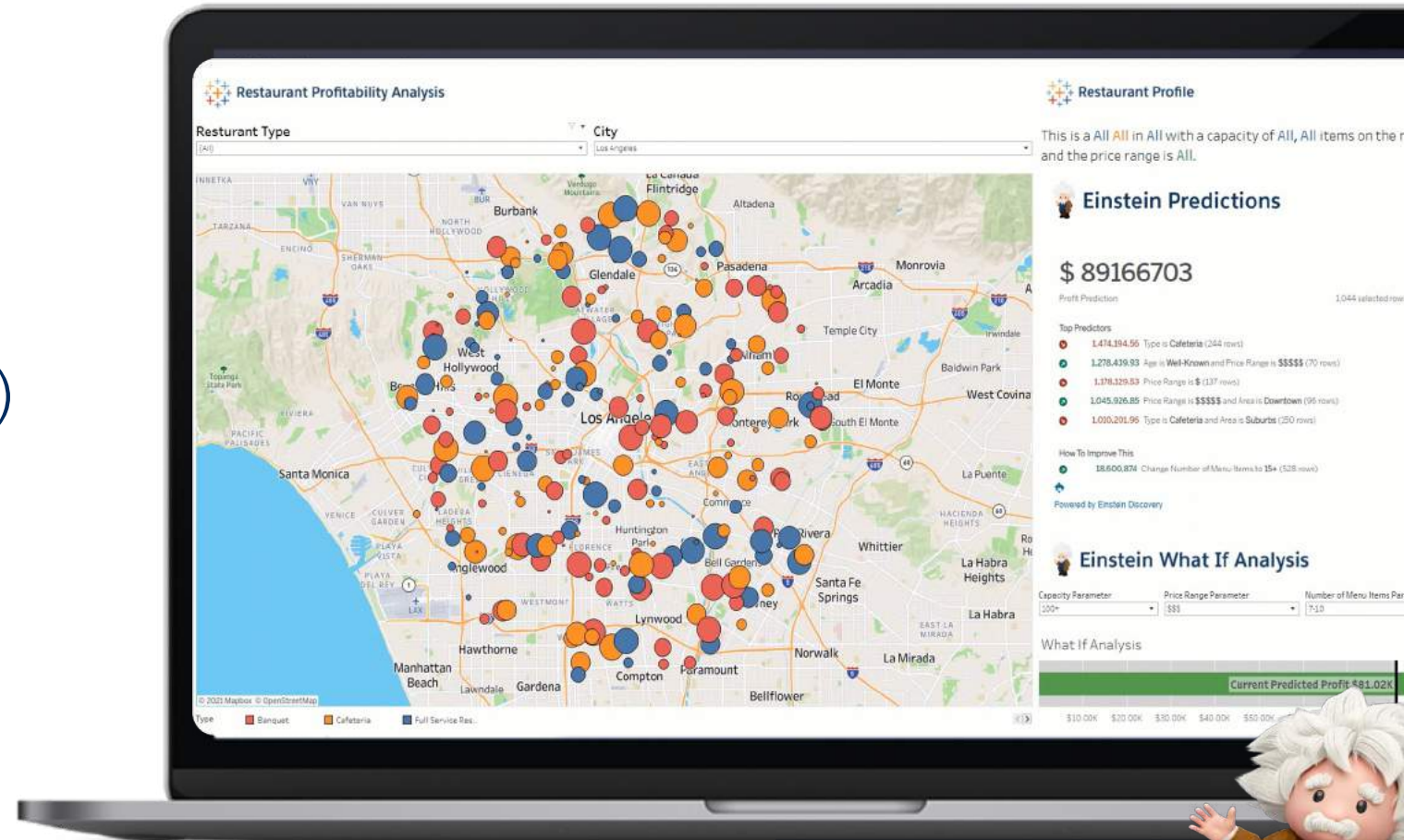
모든 비즈니스 질문의 우선순위가
똑같지는 않습니다.

정확성 우선

데이터 사이언스

테블로 비즈니스 사이언스

- 코드 없는 ML 모델 설계
- 신속하고 반복 가능한 배포
- 쉽게 이해 가능한 (no blackbox)
- 통합 및 상호 작용



아인슈타인 디스커버리 Einstein Discovery

아인슈타인 디스커버리

신뢰할 수 있고, 자동화된 모델 학습 및 신속한 배포

지도 기계 학습

코드 없는 설계

윤리적이고 투명한 ML

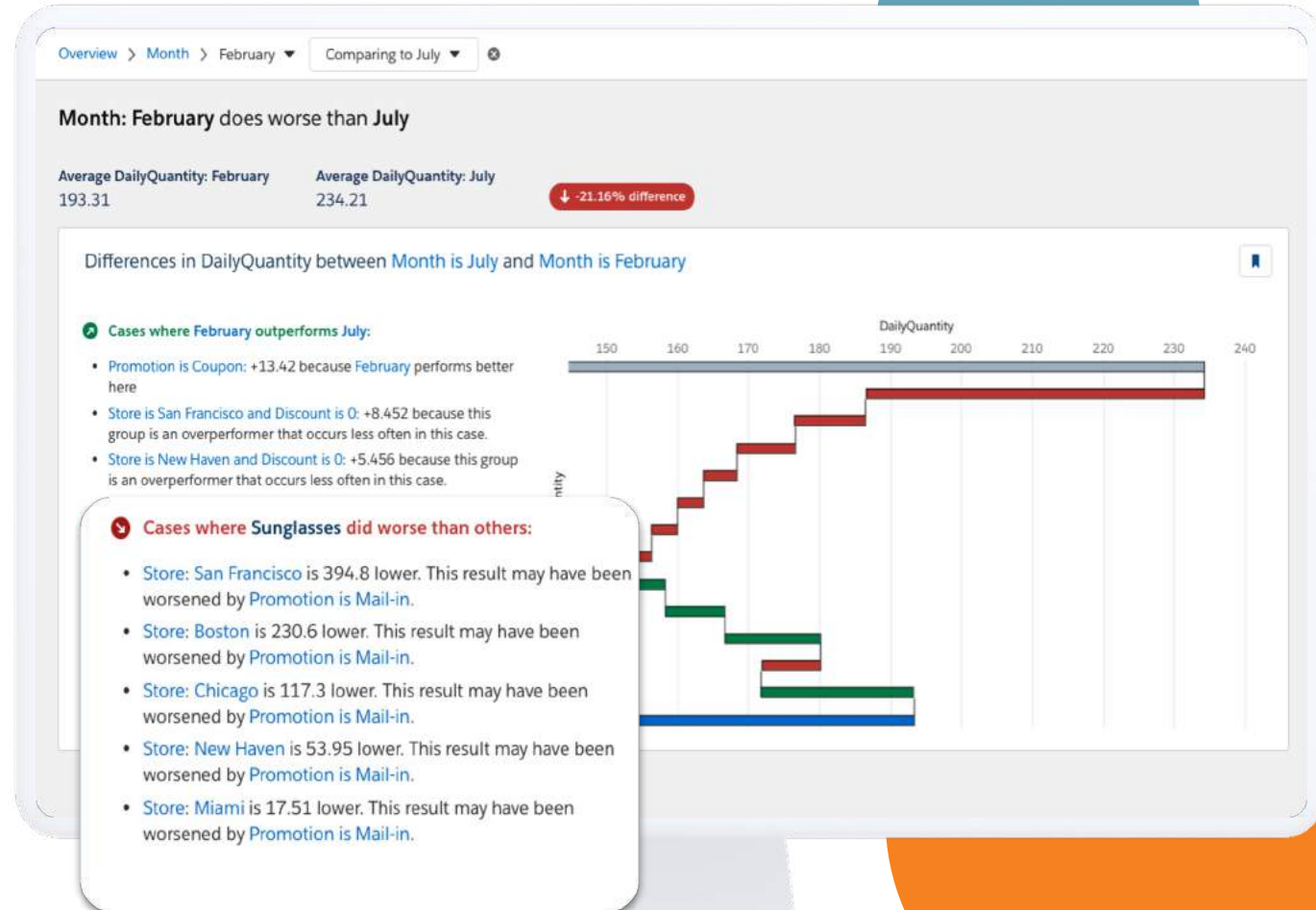
요소별 동인 분석

신속한 반복 가능

어떤 데이터도 결합 가능

실시간 모델 모니터링 가능

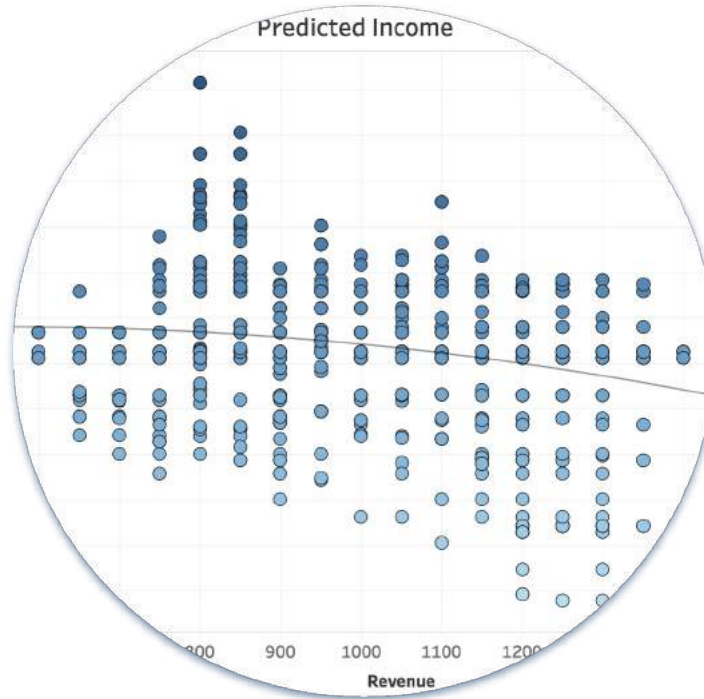
Last mile prediction delivery



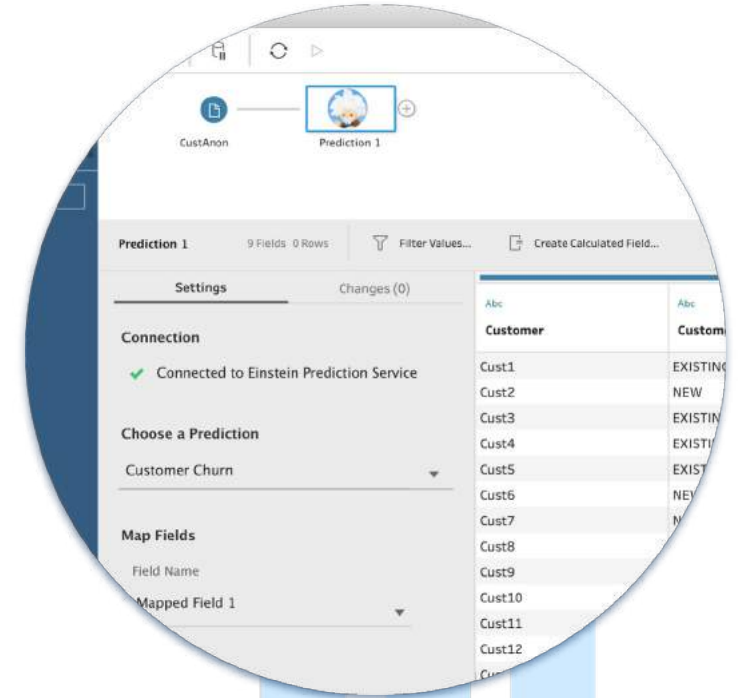
아인슈타인 디스커버리를 태블로와 통합하는 방법



대시보드 익스텐션
실시간 접근
대시보드에서 예측 결과 확인



애널리틱스 익스텐션
비주얼라이제이션에
실시간 예측 결과 통합



태블로 프렙
데이터 원본에
예측 결과 데이터 추가

데모 시나리오

구독 커머스의 이탈위험 고객 모니터링

고객ID	연령대	성별	월 구독요금	구독기간 (개월)	다음달 권장 요금제	최근 2주 이내 참여 프로모션	추천의향	감성평가	이탈여부
CX000001	31_40	남성	49000	24	인상	참여하지 않음	물어보면 추천할 것	중립	No
CX000002	41_50	남성	19000	20	인하	구독료 5% 할인 프로모션	적극 추천 의향 있음	긍정	No
CX000003	41_50	남성	19000	44	인하	구독료 5% 할인 프로모션	물어보면 추천할 것	긍정	No
CX000004	41_50	남성	29000	23	인하	구독료 5% 할인 프로모션	적극 추천 의향 있음	긍정	No
CX000005	41_50	남성	29000	4	인하	추가 사은품 증정 프로모션	적극 추천 의향 있음	중립	No
CX000006	41_50	남성	29000	27	인하	구독료 5% 할인 프로모션	적극 추천 의향 있음	긍정	No
CX000007	41_50	남성	19000	16	인하	구독료 5% 할인 프로모션	적극 추천 의향 있음	중립	No
CX000008	41_50	여성	19000	36	인하	구독료 5% 할인 프로모션	물어보면 추천할 것	긍정	No
CX000009	41_50	여성	19000	22	인상	구독료 5% 할인 프로모션	물어보면 추천할 것	부정	No
CX000010	41_50	남성	19000	34	인하	구독료 5% 할인 프로모션	추천 의향 없음	긍정	No
CX000011	41_50	여성	19000	47	인하	구독료 5% 할인 프로모션	추천 의향 없음	중립	No
CX000012	41_50	남성	19000	29	인하	구독료 5% 할인 프로모션	적극 추천 의향 있음	긍정	No
CX000013	41_50	남성	19000	30	인하	구독료 5% 할인 프로모션	적극 추천 의향 있음	중립	No
CX000014	41_50	남성	29000	12	인하	구독료 5% 할인 프로모션	물어보면 추천할 것	긍정	No
CX000015	41_50	여성	29000	32	인상	구독료 5% 할인 프로모션	물어보면 추천할 것	중립	No
CX000016	41_50	남성	19000	47	인하	추가 사은품 증정 프로모션	적극 추천 의향 있음	긍정	No
CX000017	41_50	여성	29000	35	인하	추가 사은품 증정 프로모션	물어보면 추천할 것	긍정	No

11

관심 목록

찾아보기

내가 만든 오픈서브
EinsteinBotsV2
Process Automated
Process Automated

내가 공유한 항목 Shared App User Integration User Integration

	Workforce Engagement Management	Process Automated	Process Automated	
--	---------------------------------	-------------------	-------------------	--

컬렉션 대시보드

제 목	위 치	작성 자	작성 일자	최종 수정자	최종 수정일	데이터 새로 고침
구성을 시작합니다.						

EinsteinBotsV2 EinsteinBotsV2 Process Automated Process Automated

한승세터

 Historical Adherence
  Workforce Engagement Manager
  Process Automated
  Process Automated

데이터 관리자

Intelligent Forecast Workforce Engagement Manage... Process Automated Process Automated

Long-term Omni-Channel Plan workforce Engagement Manage... Process Automated Process Automated

Omni Channel Plan Workflow Engagement Manager Process Automated Process Automated

 Omni-Channel Plan
 Workforce Engagement Management
 Process Automated
 Process Automated

기후변화 2012 보고서 2012.12.10. 10:00

제본부처: 작성일자: 최종수정일자: 최종수정일자: 데이터 새로 고침

Icon	EinsteinBotsV2	EinsteinBotsV2	Process Automated	User Integration

Historical Adherence Workforce Engagement Manager Process Automated User Integration

Historical Adherence Workforce Engagement Manage... Process Automated User Integration

Tableau Data Science



사용자의 페르소나에 따라 활용 가능한 분석 영역이 달라집니다.



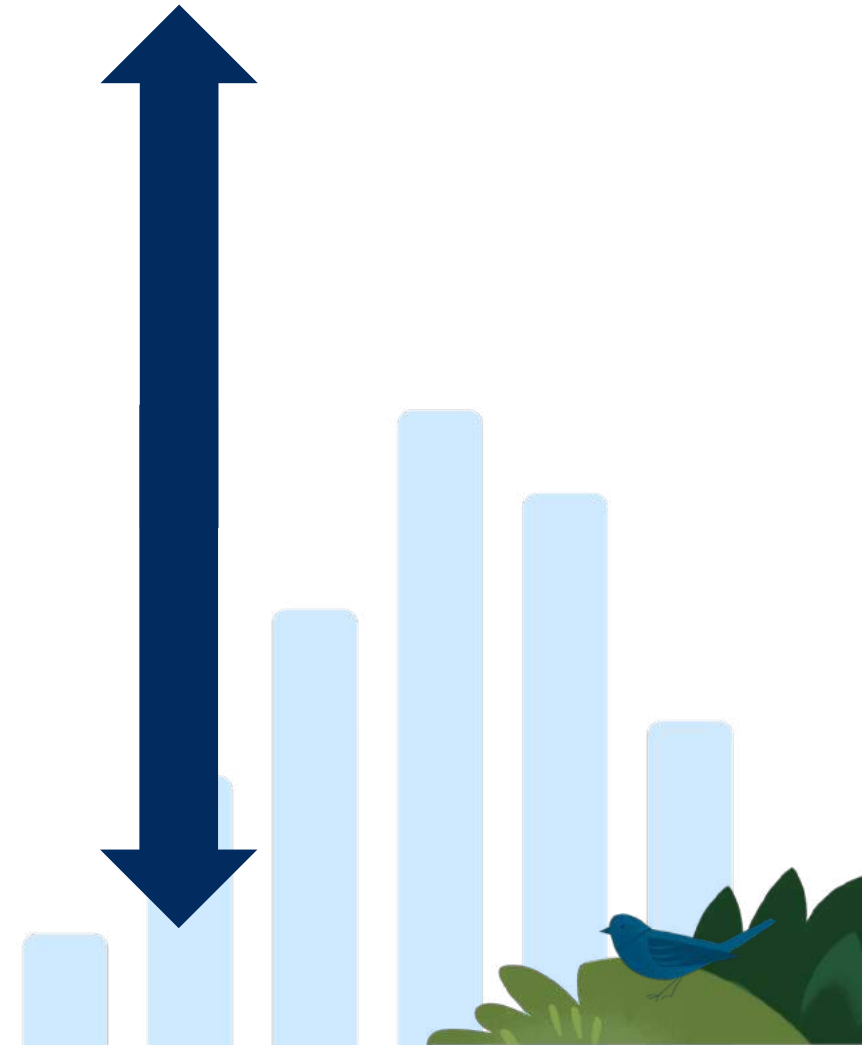
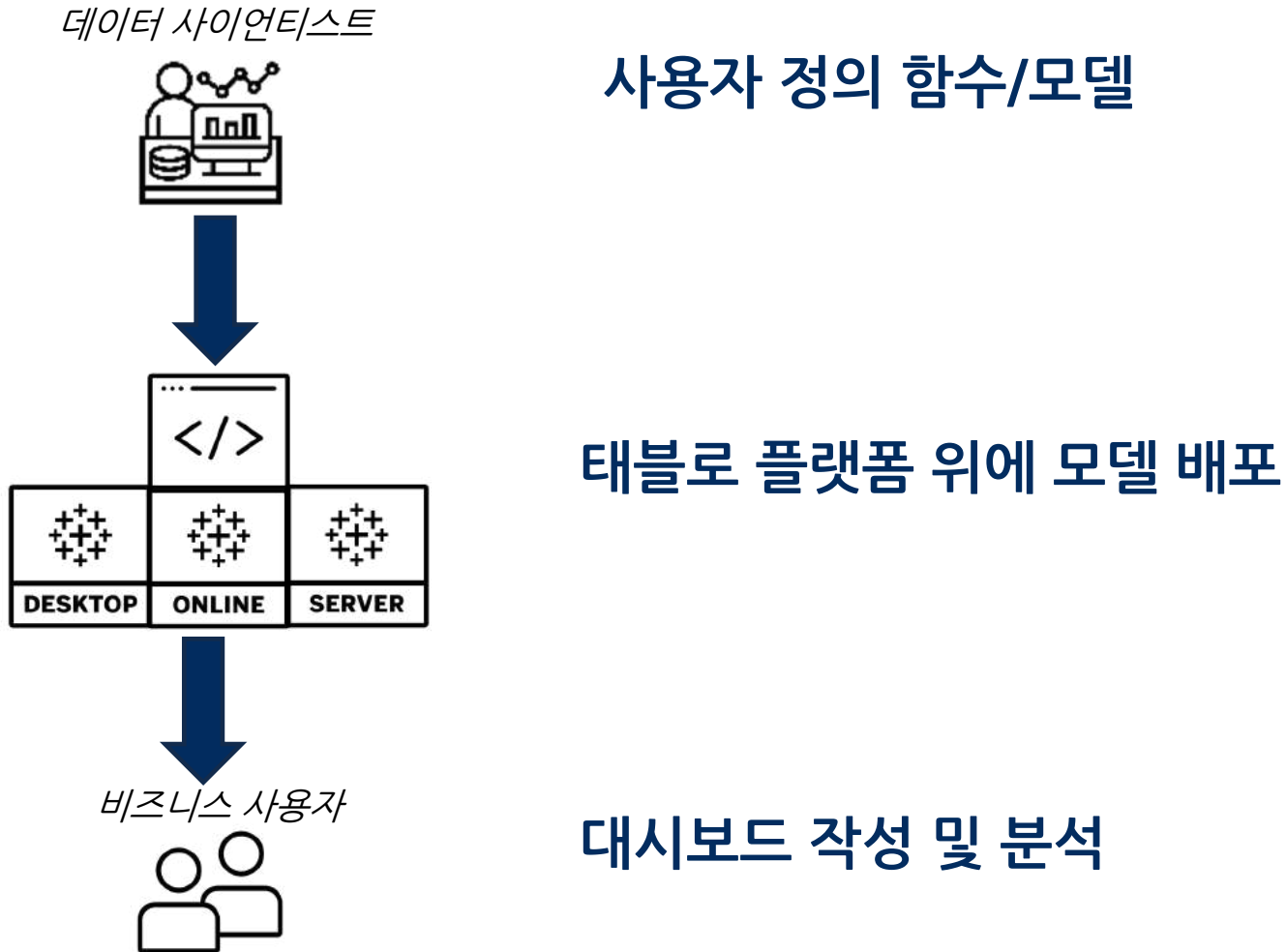
테블로 데이터 사이언스 통합

데이터 사이언스와 비즈니스 유저 간의 간극을 줄일 수 있습니다.



테블로 데이터 사이언스 통합

데이터 사이언스와 비즈니스 유저 간의 간극을 줄일 수 있습니다.



데모 시나리오

식당 입지 선정을 위한 수익 예측 시뮬레이션

Age	Area	Capacity	County Name	Number of M	Price Range	State	Type	Zip Code	Cost
Recent	Downtown	100+	Los Angeles	07-10	\$\$	California	Banquet	90002	1690000
Recent	Suburbs	15-20	Los Angeles	10-15	\$	California	Full Service R	90044	1770000
Developed	Downtown	70-90	Los Angeles	Salad & Sand	\$\$\$\$\$	California	Cafeteria	90061	1480000
Developed	Downtown	30-50	Los Angeles	05-07	\$	California	Banquet	90001	840000
Well-Known	Suburbs	100+	Los Angeles	< 5	\$\$\$\$\$	California	Banquet	90247	1630000
Well-Known	Suburbs	70-90	Los Angeles	Salad & Sand	\$\$\$\$	California	Cafeteria	90221	1140000
Well-Known	Suburbs	10-15	Los Angeles	05-07	\$\$\$\$	California	Banquet	90222	1450000
Well-Known	Downtown	70-90	Los Angeles	15+	\$\$	California	Banquet	90223	1270000
Well-Known	Suburbs	30-50	Los Angeles	05-07	\$	California	Full Service R	90224	1040000
Well-Known	Downtown	10-15	Los Angeles	< 5	\$\$\$	California	Banquet	90248	1610000
Recent	Suburbs	70-90	Los Angeles	15+	\$\$	California	Banquet	90249	2220000
Developed	Suburbs	30-50	Los Angeles	05-07	\$\$\$	California	Banquet	90251	1420000
Developed	Suburbs	50-70	Los Angeles	05-07	\$\$\$\$\$	California	Full Service R	90260	1600000
Developed	Downtown	10-15	Los Angeles	05-07	\$\$	California	Cafeteria	90262	1370000

파이썬에서 예측 함수 정의

```
Resturant_Profitability_Predictive_Model.py - /Volumes/GoogleDrive/My Drive/Works/Marketing/Datafest22/Resturant_Profitability_Predictive_Model/Resturant_Profitability_Predictive_Model.py (3.10.2)

param_grid = {}
param = [
    {'regressor': [RandomForestRegressor()],
     'regressor__n_estimators': [100, 200, 500],
     'regressor__max_depth': list(range(5, 25, 5)),
     'regressor__min_samples_split': list(range(4, 12, 2))
    },
    {'regressor': [KNeighborsRegressor()],
     'regressor__n_neighbors': [5, 10, 20, 30],
     'regressor__p': [1, 2]
    },
    {'regressor': [Lasso(max_iter=500)],
     'regressor__alpha': [0.001, 0.01, 0.1, 1, 10, 100, 1000]
    }
]
param_grid = param
clf = GridSearchCV(pipe, param_grid=param_grid,
                  cv=5, n_jobs=-1, scoring='metric')
best_clf = clf.fit(X, y)
return(best_clf.best_params_['regressor'])

#Reading Data
url= 'Restaurant_Profitability_Training_Data.csv'
df = pd.read_csv(url)

Target = 'Profit'
categorical_features = ['Area', 'Age', 'Type', 'Price Range', 'Capacity', 'Number of Menu Items']
numerical_feature = []
target = 'Profit'

label=df[target]
data= df[categorical_features+numerical_feature]

#Data Preprocessing
numeric_transformer = Pipeline(steps=[('imputer', SimpleImputer(strategy='median'))
                                     ,('scaler', StandardScaler())])
categorical_transformer = OneHotEncoder(categories='auto')

encoder = ColumnTransformer(
    transformers=[
        ('numerical', numeric_transformer, numerical_feature),
        ('categorical', categorical_transformer, categorical_features)])
encoder.fit(data)

#Model Building and Selection
clf = regressor_selection(encoder.transform(data), label, metric = 'r2')
model = clf.fit(encoder.transform(data), label)

#Wrapper function for prediction
def Profitability_Prediction(Areas, Ages, Types, Price_Ranges, Capacities, Items):
    input_data = np.column_stack([Areas, Ages, Types, Price_Ranges, Capacities, Items])
    X = pd.DataFrame(input_data, columns=['Area', 'Age', 'Type', 'Price Range', 'Capacity', 'Number of Menu Items'])
    result = model.predict(encoder.transform(X))
    return result.tolist()

#Model deployment
client = Client('http://localhost:9004/')
client.deploy('Restaurant_Profitability',
             Profitability_Prediction,
             'Returns prediction of profitability for restaurant(s).',
             override = True)
```

```
#Wrapper function for prediction
def Profitability_Prediction(Areas, Ages, Types, Price_Ranges, Capacities, Items):
    input_data = np.column_stack([Areas, Ages, Types, Price_Ranges, Capacities, Items])
    X = pd.DataFrame(input_data, columns=['Area', 'Age', 'Type', 'Price Range', 'Capacity', 'Number of Menu Items'])
    result = model.predict(encoder.transform(X))
    return result.tolist()
```

대시보드 레이아웃 <

기본값
전화

기기 미리 보기

크기

사용자 지정 크기(1400 x 827) ▼

시트

 What-if

 Overall_2 Zoom (Prediction)

개체

A

☐ 빈 페이지

다운로드

 확장 프로그램

③ 데이터에 접근

바둑판식

—III—

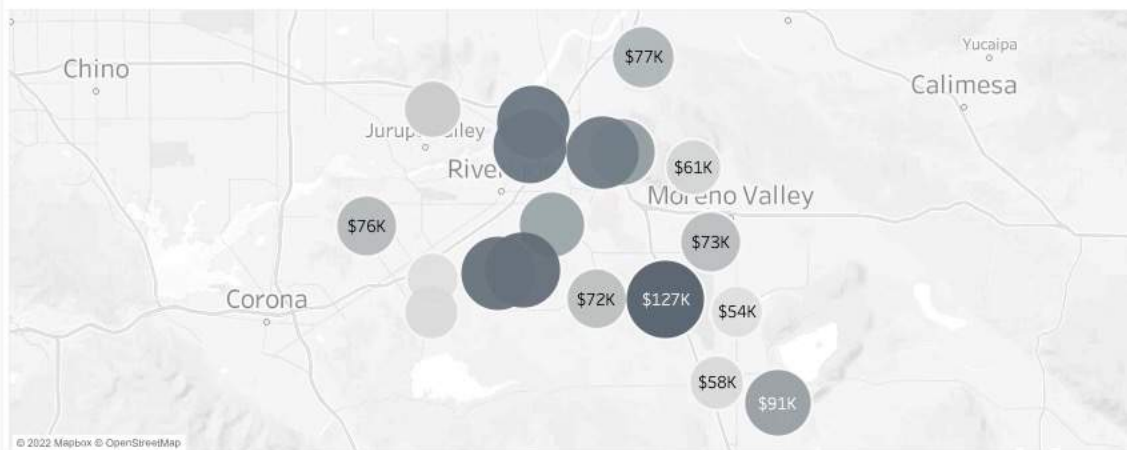
☐ 대시보드 제목 표시

식당 위치 선정 시뮬레이션 대시보드

Step 1. 관심 있는 지역을 선택하세요.



Step 2. 선택한 지역에서 매입 가능한 건물 위치와 예상 수익을 확인하세요.



Step 3. 예상 수익 vs What-if 수익을 비교해보세요.

선택된 지역: 92507

Age

Recent

Number of Items

7-10

Area

Downtown

Capacity

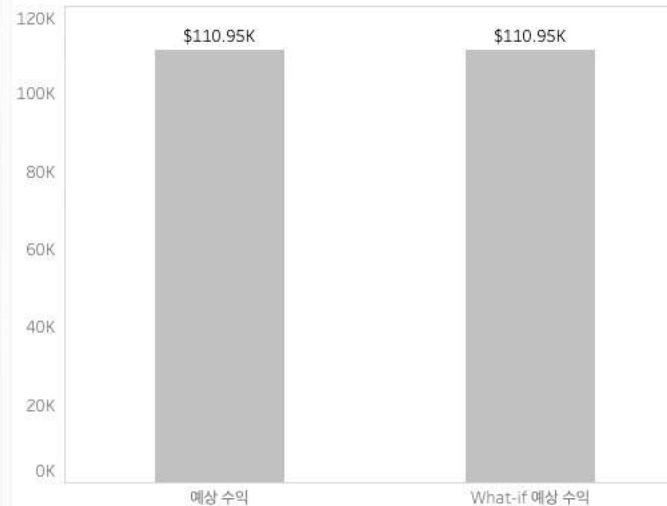
20-30

Price Range

\$\$\$

Type

Banquet



사용자의 페르소나에 따라 활용 가능한 분석 영역이 달라집니다.

비즈니스 사이언티스트
가이드드 & 자동화 ML

No-code ML Model Authoring

Simulation



Scoring in Data Pipelines



Prediction Monitoring

Visual Advanced Analytics &
Predictive Calculations



BI 분석가 & 익스플로러
AI-증강 분석

Discovery Stories
Explain Data-
Report Insights



Ask Data
Conversational Queries



End Users
Better, Faster
Decision-making with AI

데이터 사이언티스트
사용자정의 모델 구축 및 배포



BYOM



Analytics Extensions



Notebooking



Prep Script Nodes

Thank you

