

데이터베이스(002) 기말 프로젝트 발표

은행 정기예금 마케팅 데이터 분석을 통한 마케팅 대상 추천 시스템 구축

Table of Contents

01 프로젝트 개요

02 데이터베이스 설계

03 SQL 분석 및 시각화

04 프로그램 구현

05 마무리

01 프로젝트 개요

주제 선정 이유

- 은행은 고객을 대상으로 다양한 금융 상품 마케팅을 진행하지만, 실제 정기 예금 신규 가입률은 낮은 편이다.
- 불특정 다수에게 일괄적으로 마케팅을 실시할 경우, 불필요한 비용이 발생하고 효율성도 떨어진다.
- 데이터 기반으로 고객 특성을 분석하고, 가입 가능성이 높은 대상을 선별해 마케팅 효율을 높일 수 있는 시스템의 필요성을 느꼈다.

프로젝트 목적

- 은행 정기예금 마케팅 데이터를 체계적으로 분석하고, SQL 기반으로 정규화 된 데이터베이스를 구축하여 효율적으로 데이터를 관리하는 것
- 더 나아가, 분석 결과를 바탕으로 실제 마케팅 대상 추천 프로그램 구현
- 이를 통해 데이터 기반 의사결정의 실제 과정을 경험하고, 실질적인 마케팅 전략 수립에 도움이 되는 인사이트를 도출하고자 하였다.

활용 데이터

- UCI Machine Learning Repository에서 제공하는 Bank Marketing 데이터셋
- 약 4만 5천 명의 은행 고객 정보를 포함하고 있으며, 고객의 나이, 직업, 결혼 여부 등 다양한 변수를 담고 있다.
- 데이터는 CSV 파일 형태로 제공되며, 범주형, 수치형, 이진형 변수가 혼합되어 있다.

02 데이터베이스 설계 - 원본 CSV

피처명	설명	데이터 타입	예시
age	고객 나이	정수	30
job	직업	문자열(범주형)	'admin.'
marital	결혼 여부	문자열(범주형)	'married'
education	교육 수준	문자열(범주형)	'secondary'
default	신용불량 여부	'yes' / 'no'	'no'
balance	은행 잔액	정수	1500
housing	주택 대출 여부	'yes' / 'no'	'yes'
loan	개인 대출 여부	'yes' / 'no'	'no'
contact	연락 수단	문자열(범주형)	'cellular'
day	마지막 연락 일	정수	5
month	마지막 연락 월	문자열	'may'
duration	통화 시간(초)	정수	120
campaign	캠페인 동안 연락 횟수	정수	2
pdays	이전 캠페인 이후 경과 일수	정수	999
previous	과거 캠페인 접촉 횟수	정수	0
poutcome	이전 캠페인 결과	문자열(범주형)	'failure'
y	정기예금 가입 여부	'yes' / 'no'	'no'

02 데이터베이스 설계 - 정규화

1) 1차 정규화 (1NF) : 반복 속성 제거

- 각 셀이 단일 값을 갖도록 구조 개선
- 원본 데이터에서 직접적으로 반복되는 속성은 없었지만, 범주형 속성들이 문자열 형태로 저장되어 있었음
- 이를 해결하기 위해 각 범주형 속성을 별도의 코드 테이블로 분리!

2) 2차 정규화 (2NF) : 부분 종속성 제거

- 초기에는 모든 정보가 하나의 테이블에 포함되어 있어, 고객의 기본 정보가 캠페인마다 중복되어 저장되는 문제 발생
- 이를 해결하기 위해 고객 정보를 담은 Customer 테이블과 캠페인 기록을 담은 Campaign 테이블로 분리!
- 그 결과, 한 고객이 여러 번의 마케팅 캠페인에 참여한 경우에도 고객 정보의 중복 없이 효율적으로 관리할 수 있게 됨

3) 3차 정규화 (3NF) : 이행적 종속성 제거

- education, contact, poutcome과 같은 속성들은 고객이나 캠페인에 종속되지만, 그 자체로도 의미 있는 독립적인 속성이므로 각각을 별도의 코드 테이블로 분리!
- contact와 poutcome도 동일한 방식으로 처리하여, 코드 관리의 일관성을 확보하고 향후 새로운 범주가 추가될 때의 확장성도 고려함

02 데이터베이스 설계 - 정규화

4) 주요 제약조건

- 각 테이블의 기본키는 AUTO_INCREMENT 속성을 부여하여 자동으로 증가하는 정수값을 사용하도록 함
→ 새로운 데이터가 삽입될 때마다 고유한 식별자가 자동으로 생성되어, 중복이나 식별자 관리 문제 방지
- Customer 테이블과 Campaign 테이블 간에는 customer_id를 통해 연결하고, 각 범주형 속성 테이블들과는 해당하는 ID 컬럼을 외래키로 설정
→ 존재하지 않는 코드값이 입력되는 것을 방지하고, 데이터 간의 일관성 유지
- 각 코드 테이블의 name 컬럼에는 UNIQUE 제약조건을 설정하여 동일한 값이 중복으로 저장되는 것을 방지함
→ 'admin.', 'management' 등의 직업 코드가 중복되어 저장되지 않도록 관리

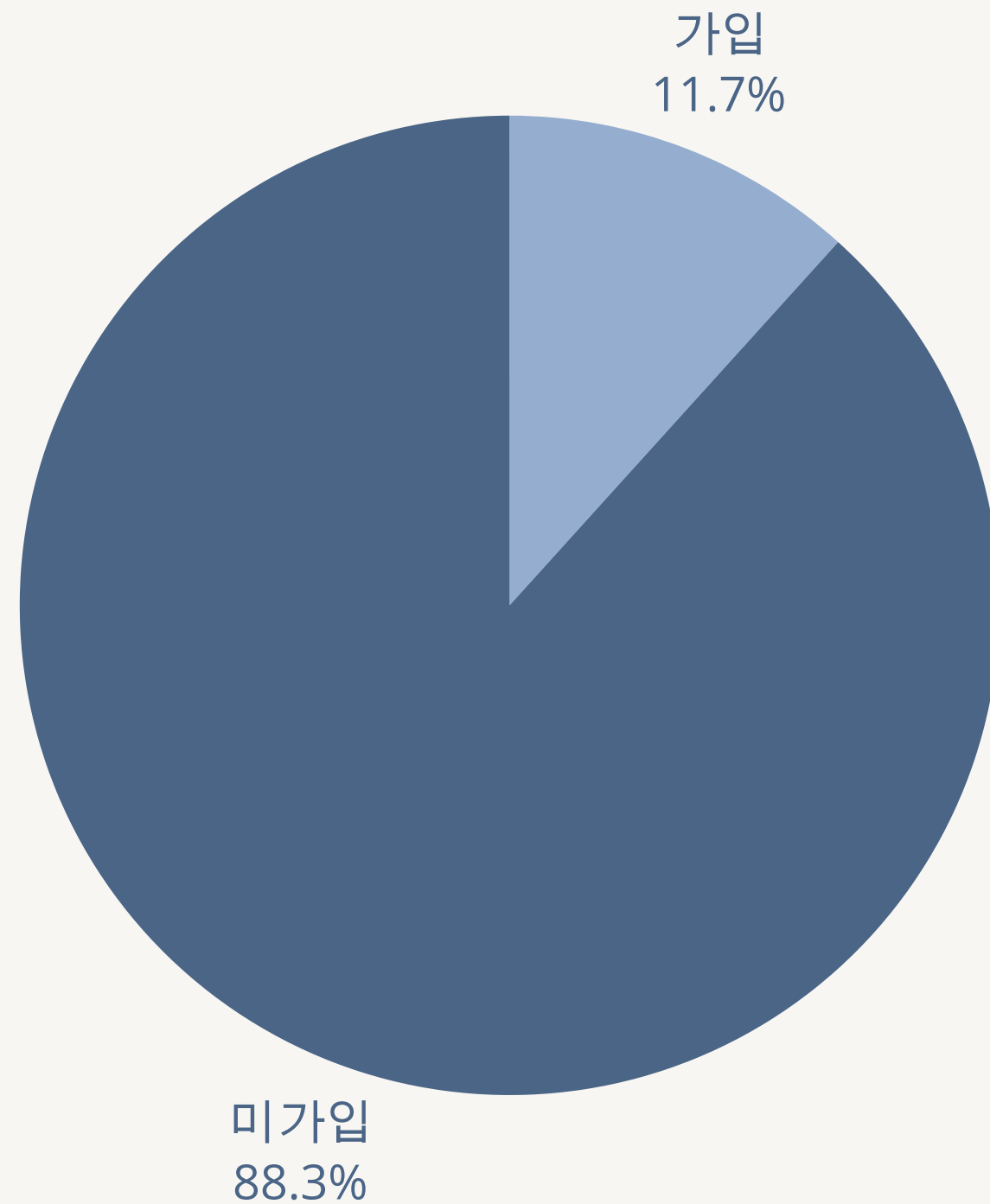
02 데이터베이스 설계 - 정규화

5) 데이터 삽입 방법

- pandas & pymysql 라이브러리 활용 → 효율적으로 데이터 전처리 및 데이터베이스 구축
- INSERT IGNORE 구문을 사용하여 동일한 값이 중복으로 삽입되는 것을 방지함
- 결측값 처리 - 원본 데이터에서 빈 값이나 null 값이 발견되는 경우, 일관되게 'unknown'으로 대체하여 처리
→ 데이터 분석 과정에서 결측값으로 인한 오류를 방지하고, 향후 분석에서도 일관된 기준을 적용할 수 있었음
- 데이터 삽입 순서는 참조 무결성을 고려하여 부모 테이블부터 자식 테이블 순으로 진행!
→ 먼저 각 코드 테이블(Job, Education 등)에 데이터를 삽입한 후, Customer 테이블에 고객 정보를 저장하고, 마지막으로 Campaign 테이블에 캠페인 이력을 저장함
- 삽입 과정에서 pandas의 to_sql() 메서드를 활용하여 대량의 데이터를 한 번에 효율적으로 삽입함

03 SQL 분석 및 시각화

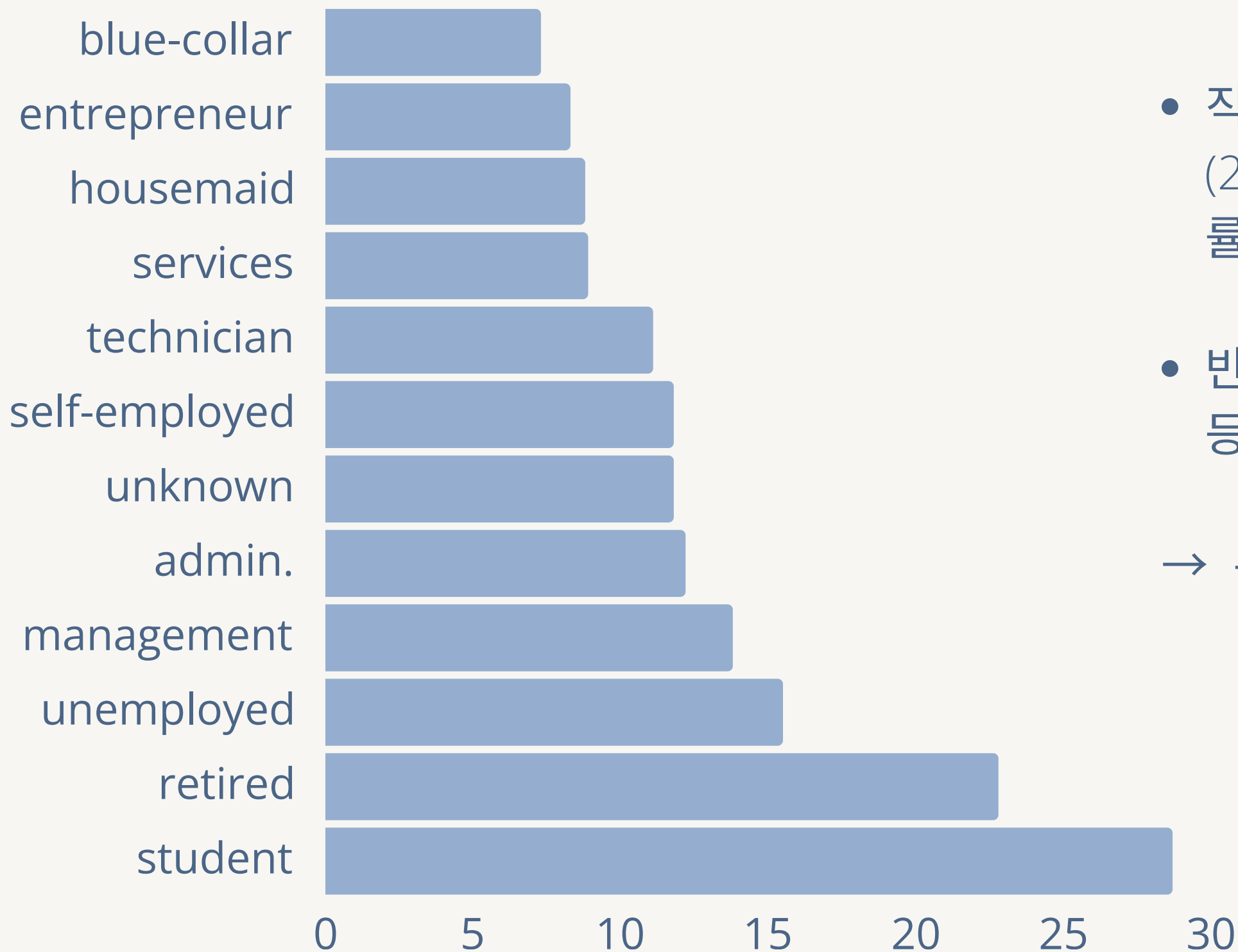
1) 전체 고객 가입률



- 전체 고객 수는 약 45,211명이며, 이 중 5,289명이 정기예금에 가입해 가입률은 약 11.7%로 나타남
- 이는 실제 마케팅의 성공률이 매우 낮다는 점을 보여줌

03 SQL 분석 및 시각화

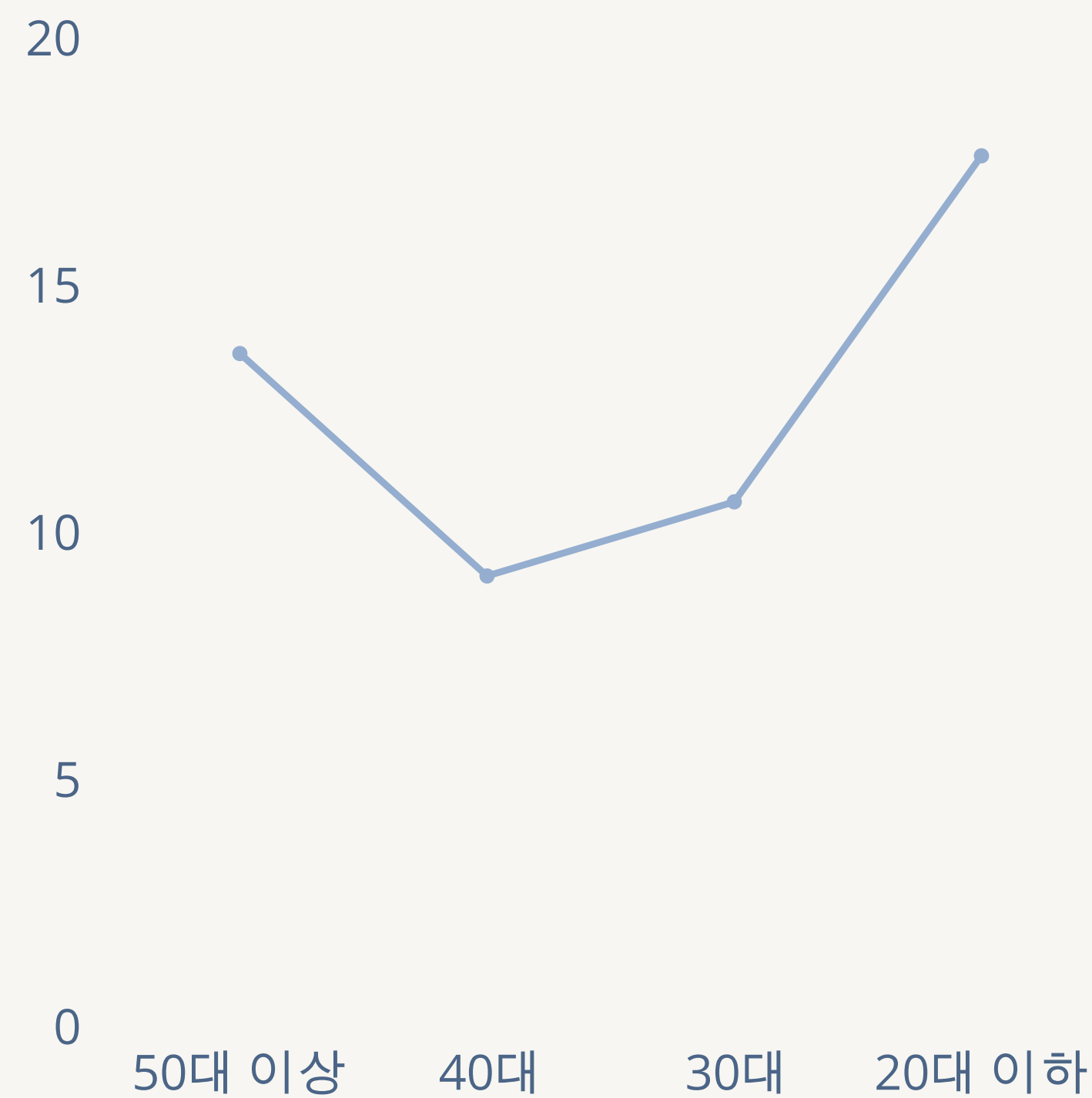
2) 직업별 가입률



- 직업별로 가입률을 분석한 결과, 학생(28.7%), 은퇴자(22.8%), 실업자(15.5%), 관리자(13.8%) 순으로 가입률이 높았음
 - 반면, 블루칼라(7.3%), 자영업(8.3%), 서비스직(8.9%) 등은 가입률이 낮은 편이었음
- 특정 직업군이 정기예금 상품에 더 관심이 많거나, 마케팅에 더 잘 반응!

03 SQL 분석 및 시각화

3) 연령대별 가입률



- 연령대별로는 50대 이상(13.6%)과 20대 이하(17.6%)에서 가입률이 높게 나타남

03 SQL 분석 및 시각화

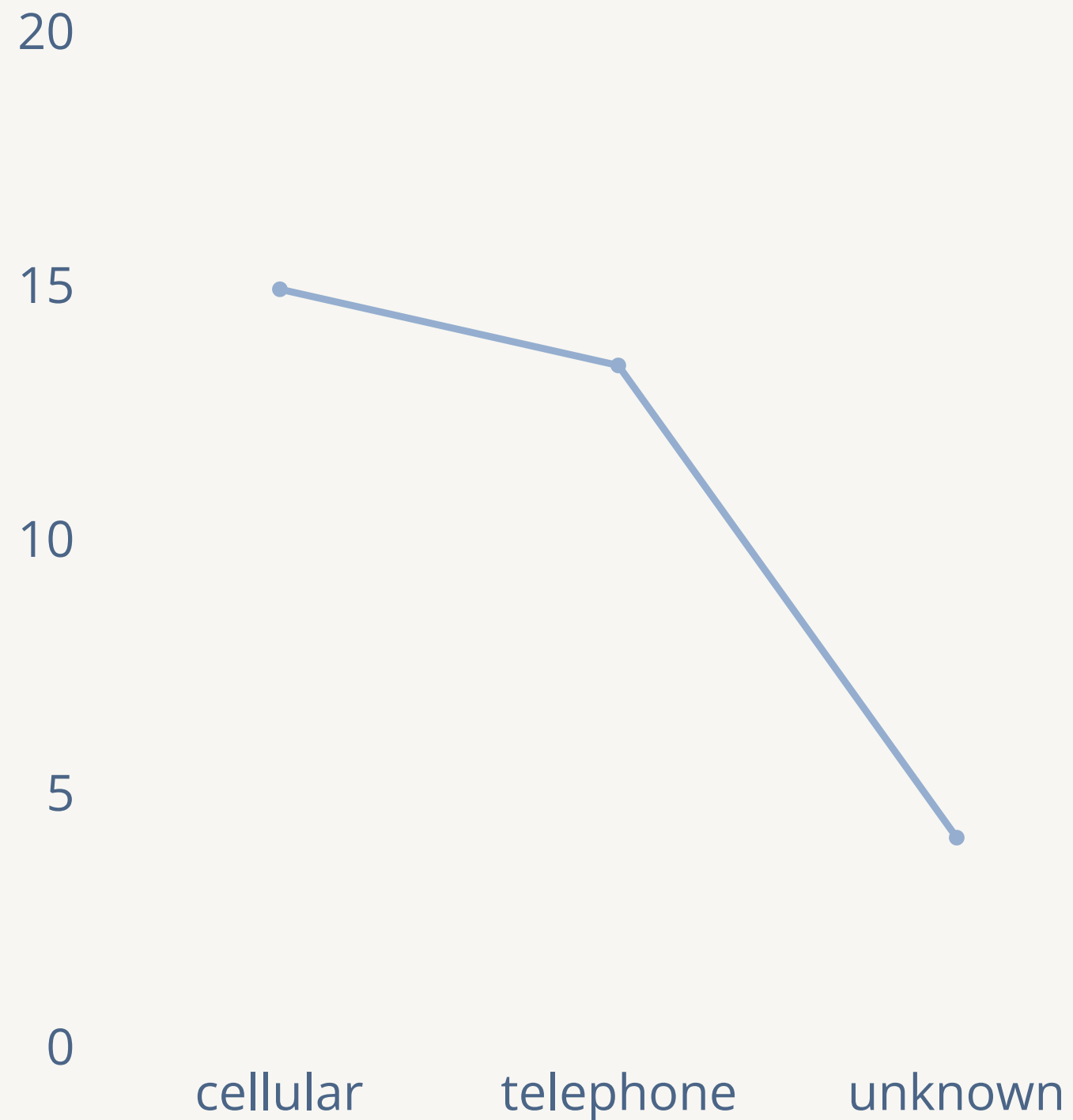
4) 월별 캠페인 성공률 추이



- 월별로 분석한 결과, 3월(52.0%), 9월(46.5%), 10월(43.8%), 12월(46.7%) 등 특정 시기에 가입률이 급상승하는 경향을 보임

03 SQL 분석 및 시각화

5) 통화 수단별 가입률



- 연락 방식별로는 휴대전화(cellular)를 통한 연락 시 가입률이 14.9%로 가장 높았고, 전화(telephone) 13.4%, 미확인(unknown) 4.1%이었음
- 이는 휴대전화가 가장 효과적인 마케팅 채널임을 보여줌

03 SQL 분석 및 시각화

6) 이전 캠페인 영향 분석



- 이전 캠페인에 성공한 경험이 있는 고객의 경우, 재가입률이 64.7%에 달해 매우 높게 나타남
- 반면, 이전 캠페인에서 실패하거나 결과가 미확인인 경우에는 가입률이 각각 12.6%, 9.2%로 낮게 나타남
- 이는 과거 마케팅 경험이 향후 행동에 큰 영향을 미친다는 점을 의미!

04 프로그램 구현

1) SQL 분석 결과를 반영한 점수화 규칙

- 직업:
학생/은퇴자: +3점
실업자/관리자/경영직: +2점
자영업/미확인/기술직: +1 점
- 연령대:
20대 이하 또는 50대 이상: +2점
30~39세: +1 점
- 연락 월:
3월, 9월, 10월, 12월: +2점
- 연락 방식:
휴대전화(cellular): +2점
전화(telephone): +1 점
- 이전 캠페인 결과:
성공: +4점 (64.7% 재가입률)
실패: -1 점

2) 점수 범위에 따른 추천 등급

점수 범위	추천 등급	설명
8점 이상	적극 추천	가입 가능성 매우 높음
5~7점	추천	가입 가능성 높음
3~4점	보통	신중히 고려
2점 이하	비추천	가입 가능성 낮음

04 프로그램 구현

3) python streamlit 라이브러리를 활용한 웹 기반 인터페이스 구현

정기예금 마케팅 대상 추천

나이

22

-

+

연락 월

jun

▼

직업

student

▼

연락 방식

cellular

▼

이전 캠페인 결과

☐ success

☐ failure

☒ unknown

가입 가능성 분석

추천 (가입 가능성 높음)

05 마무리

- 이번 프로젝트를 통해 데이터베이스 설계, SQL 분석, Python 프로그래밍, 그리고 실제 마케팅 전략 수립까지 데이터사이언스의 전 과정을 경험할 수 있었다.
- 단순한 통계 분석을 넘어, 데이터베이스를 직접 구축하고, SQL로 데이터를 가공/분석하는 과정에서 데이터의 구조화와 실용성의 중요성을 다시 한번 느꼈다.
- 분석 결과를 실제 추천 시스템 프로그램으로 구현함으로써, 데이터 기반 의사결정의 실질적인 효과와 한계를 모두 체감할 수 있었다.
- 향후에는 머신러닝 기반의 예측 모델까지 확장해, 더욱 정교한 마케팅 대상 추천 시스템을 개발해보고 싶다.

감사합니다
