

파일럿 프로젝트 결과 보고서

<연체 현황 파악>

(전북은행)한혜형
(광주은행)김은민

데이터 정의 및 미션

Sheet	column	column설명	비고
bank_data	acno	계좌번호	
	sex	성별	
	brth_dt	생년월일	
	age	나이	
	exec_dt	대출실행일	2019~2021년
	exec_amt	취급액	
	cur_bal	잔액	
	ovrd_dvcd	연체여부	연체 : 1 / 정상 : 0
cb_data	acno	계좌번호	
	CB	CB등급	
	SP	SP등급	
	R	R등급	
	PI	PI등급	

Mission	Mission명
Mission 1)	√ bank_data에 연체금액을 계산하여 'ovrd_amt'라는 칼럼을 만드세요.
Mission 2)	√ bank_data와 cb_data를 결합시키세요.
Mission 3)	√ 결합된 data를 토대로 각 외부 CB등급(4개)별 연체율을 계산하세요.
Mission 4)	√ 계산된 연체율을 토대로 subplot을 이용해 연체율을 그려주세요.(꺾은선)

데이터 전처리

1. 연체금액 산출

* 연체여부 컬럼 활용

```
df_bank['delay_amt'] = [df_bank.cur_bal[i] if df_bank.ovrd_dvcd[i] == 1 else 0  
                        for i in range(len(df_bank))]
```

2. 데이터 결합

```
df_mg = pd.merge(df_bank, df_cb, on = 'acno')  
df_mg.head()
```

데이터 전처리

3. 분기 컬럼 생성

* make_quarter 라는 함수 생성 후 대출실행일 컬럼에 적용

	acno	sex	brth_dt	age	exec_dt	exec_amt	cur_bal	ovrd_dvcd	dlay_amt	CB	SP	R	PI	exec_yq
0	37201901102569	F	19760424	46	20190101	25000000	9684729	0	0	8	6	3	5	2019.1Q
1	37201901114919	M	19610205	61	20211004	20000000	1911878	0	0	1	8	4	2	2021.4Q
2	37201901172631	M	19650714	57	20190123	25000000	14145389	0	0	10	9	9	1	2019.1Q
3	37202001038750	F	19560524	66	20190709	5000000	0	0	0	2	8	2	5	2019.3Q
4	37201901092321	F	19690709	53	20200516	30000000	4752273	0	0	7	5	8	3	2020.2Q

4. 신용등급별 연체율 생성

* make_dlay_rate(df_mg, 'R')
-> 데이터프레임과 CB사 입력

re_rating	1-2	3-4	5-6	7-8	9-10
exec_yq					
2019.1Q	0.000000	3.008064	1.750636	4.616566	12.520980
2019.2Q	0.149828	2.620736	3.227408	7.684074	13.998793
2019.3Q	0.000000	2.065129	3.752618	10.175667	14.820081
2019.4Q	0.000000	4.442823	10.565735	6.938318	16.986671
2020.1Q	1.360942	2.867876	8.879818	10.868152	10.858211

신용등급

데이터 전처리

5. 시각화 함수 구현

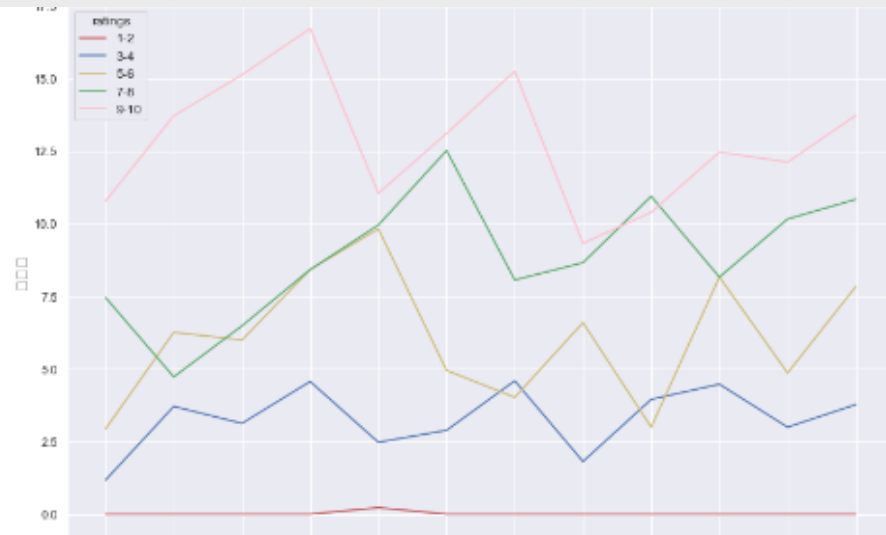
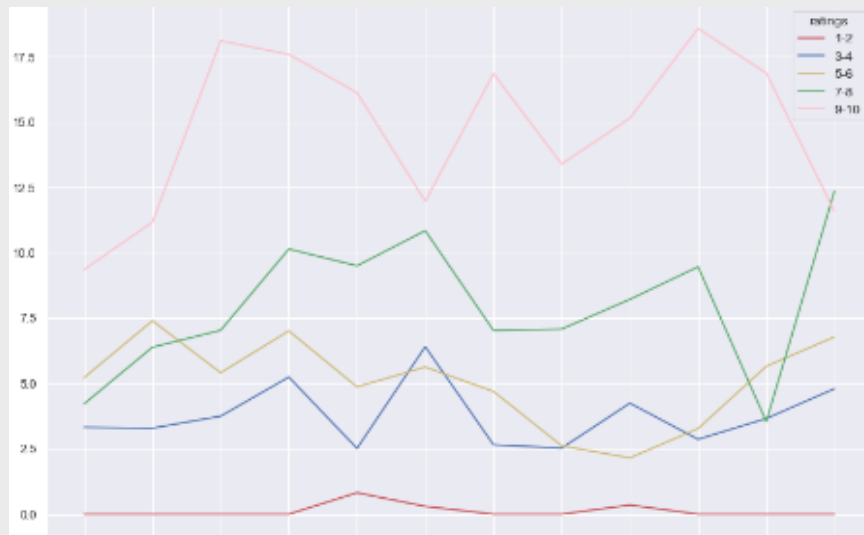
* param = '신용평가사' 입력 시 그에 맞는 그래프 호출

```
def draw_data(param):  
    sns.set()  
  
    if param == 'CB':  
        df_cb = make_dlay_rate(df_mg, 'CB')  
  
    elif param == 'SP':  
        df_sp = make_dlay_rate(df_mg, 'SP')
```

데이터 시각화

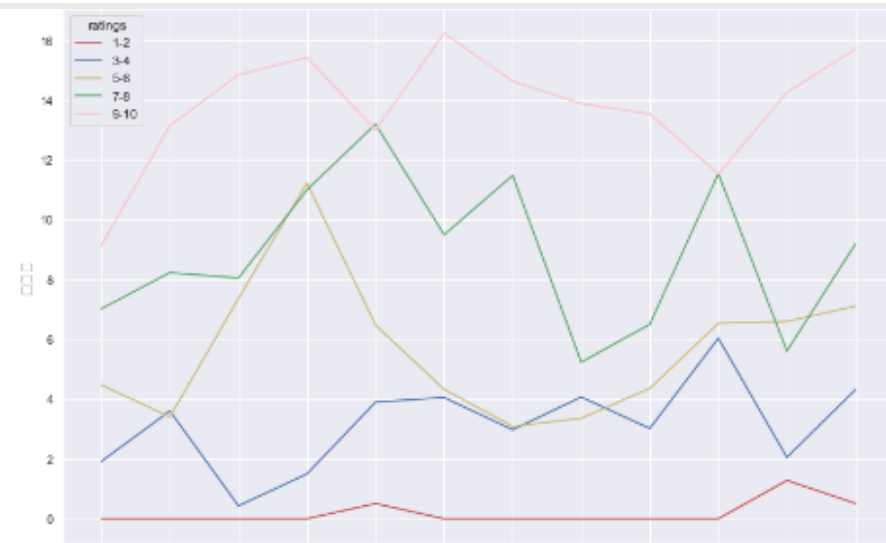
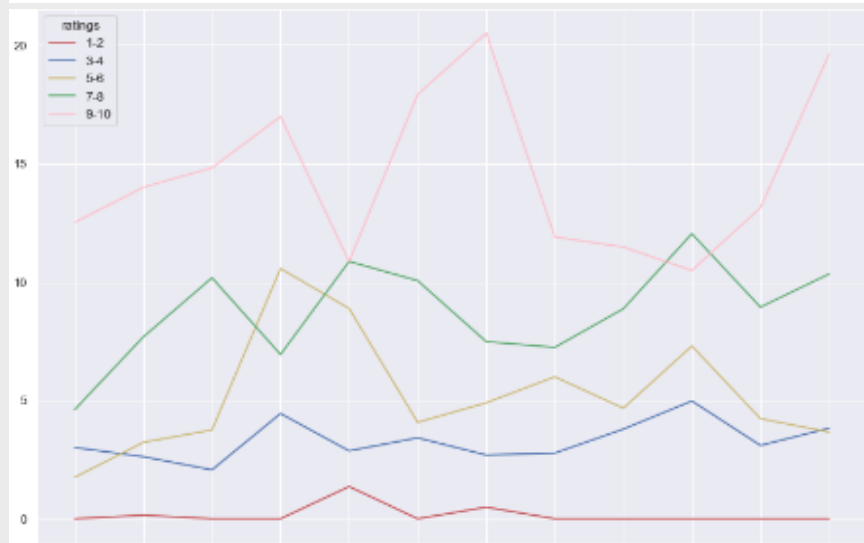
<참고> x축 : 2019~2021년 4분기, y축 : 연체율

CB



SP

R



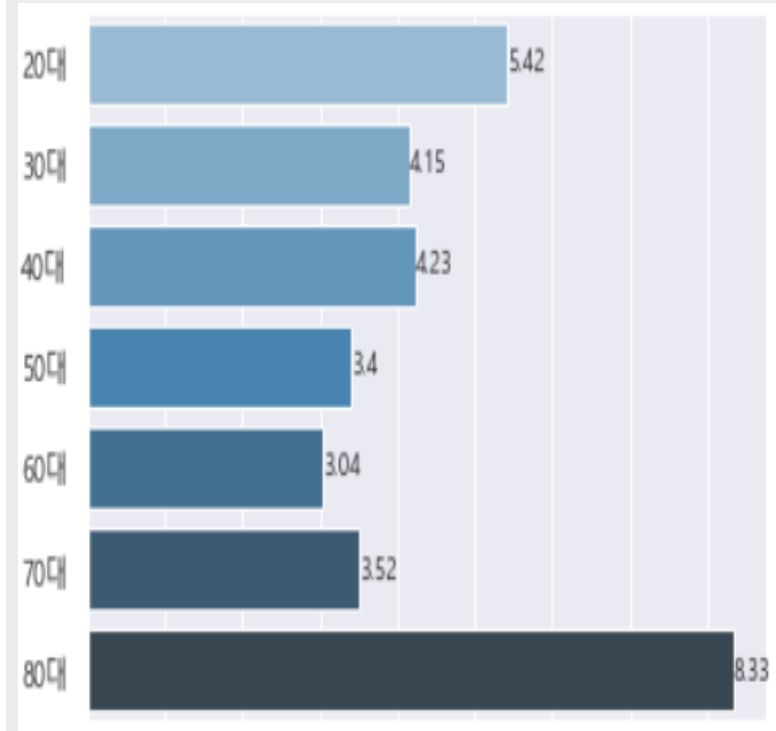
PI

* 해석 : (모든 CB사에서) 등급이 낮을수록 연체율이 높으며, 연체율의 변동성도 크다는 공통점을 보인다.

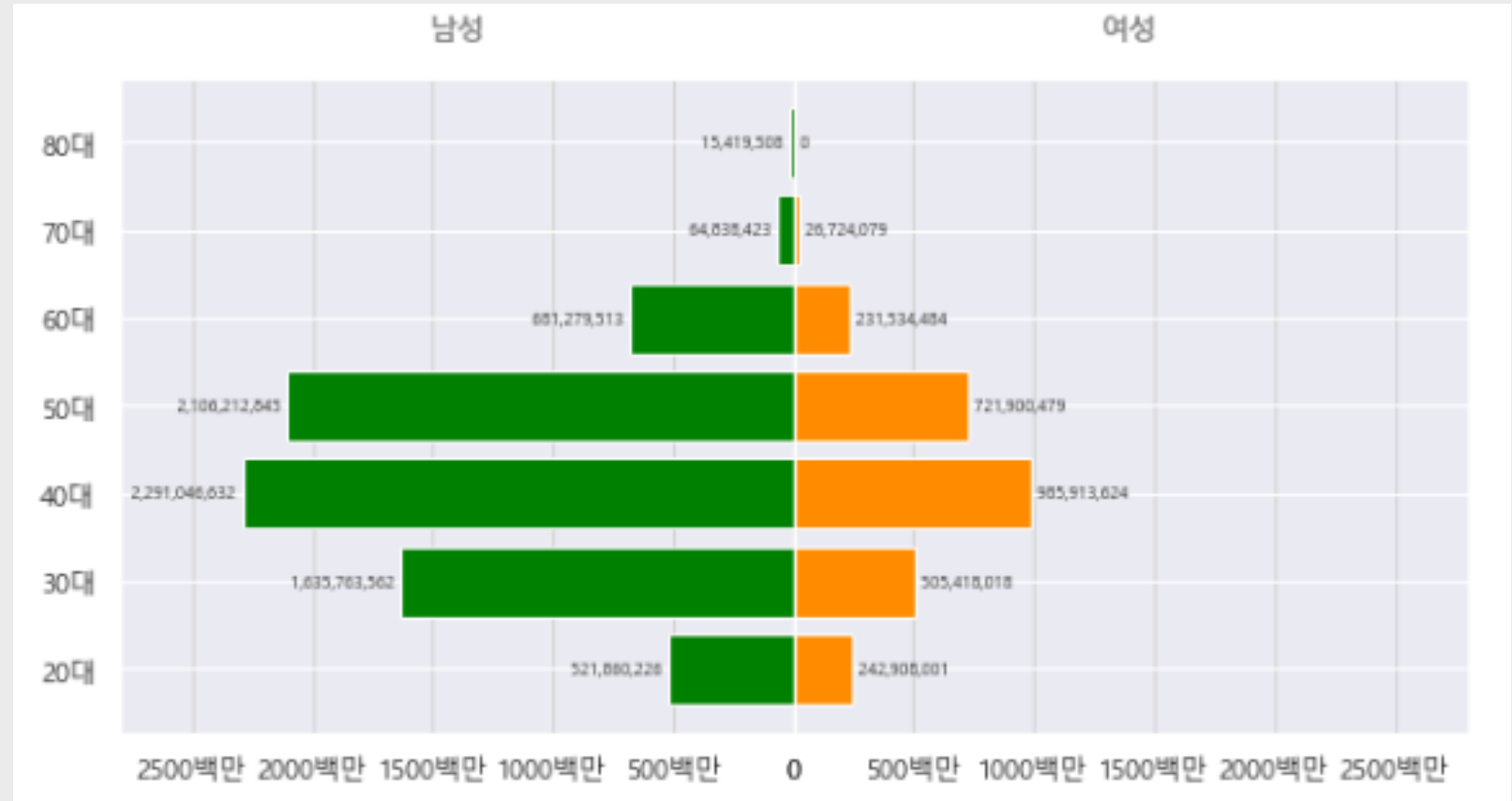
데이터 시각화

<참고> x축 : 2019~2021년 4분기, y축 : 연체율

1. 연령대별 연체여부



1. 연령대 및 성별 연체금액



배운점

1. 요구사항 정리 단계의 중요성

- 대출, 연체관련 업무의 이해도가 부족한 상태에서 이해했다고 착각하고 기획 및 분석 시작함
- 분석 작업 진행 중에 의뢰자에게 확인 후 재 작업을 진행함

2. 단계별 결과 데이터 확인의 중요성

- 데이터 분석 시 구현한 로직이 실행 되었을 때 일부 데이터만 확인 후 진행한 경우가 있었음
- Key값을 기준으로 left join한 경우 작업 전과 작업 후 raw데이터 개수가 늘어난 점을 나중에 확인 하여 재 작업을 진행함