

진단평가(데이터분석,시각화)

2024년 6월 기준 전국 여행업체 공공 데이터를 기반으로 분석에 필요한 전처리 및 조회, 추출, 시각화 작업을 수행하시오.
(제공해드린 “travel.csv” 파일을 c“\python_work 폴더에 복사한 후, 진행하세요.)

1. “travel.csv” 파일을 읽어서 데이터프레임 데이터로 변환하여 가져오시오.
데이터프레임 변수는 df로 한다. (배점 :5점)

2. df 데이터프레임 데이터 자료형 구성을 조회하시오. (배점 :5점)

```
개방자치코드      int64
인허가일자        object
영업상태코드      int64
상세영업상태      object
소재지            object
도로명            object
사업장명          object
자본금            int64
dtype: object
```

3. 결측치 값을 조회하시오. (배점 :5점)

```
개방자치코드      0
인허가일자        0
영업상태코드      0
상세영업상태      3
소재지            0
도로명            883
사업장명          0
자본금            0
dtype: int64
```

4. ① 중복 데이터 개수를 조회하시오.
3

② 중복 데이터가 존재하고 있으므로, 중복된 데이터도 조회하시오. (배점 :5점)

	개방자치코드	인허가일자	영업상태코드	상세영업상태	소재지	도로명	사업장명	자본금
2480	3170000	2010-03-16	1	영업중	서울특별시 금천구 가산동 월드메르디앙 1차 1414호	서울특별시 금천구 벚꽃로 254, 1414호 (가산동, 월드메르디앙1차)	아이빛연구소(주)	34591000
2481	3170000	2010-03-16	1	영업중	서울특별시 금천구 가산동 월드메르디앙 1차 1414호	서울특별시 금천구 벚꽃로 254, 1414호 (가산동, 월드메르디앙1차)	아이빛연구소(주)	34591000
5007	3330000	2015-07-02	1	영업중	부산광역시 해운대구 좌동 1427-2 부산좌동 창업지원주택	부산광역시 해운대구 좌동로 166, 101동 1303호 (좌동, 부산좌동 창업지원주택)	요컴퍼니	35043000
5008	3330000	2015-07-02	1	영업중	부산광역시 해운대구 좌동 1427-2 부산좌동 창업지원주택	부산광역시 해운대구 좌동로 166, 101동 1303호 (좌동, 부산좌동 창업지원주택)	요컴퍼니	35043000
16279	6520000	2016-07-21	1	영업중	제주특별자치도 서귀포시 대포동 760-3번지	제주특별자치도 서귀포시 대포로 26 (대포동)	(주)디스커버 제주	31754000
16280	6520000	2016-07-21	1	영업중	제주특별자치도 서귀포시 대포동 760-3번지	제주특별자치도 서귀포시 대포로 26 (대포동)	(주)디스커버 제주	31754000

5. 결측치가 있는 ‘상세영업상태’와 ‘도로명’ 열(column)의 결측치를 삭제하시오. (배점 :5점)

개발자치코드	0	개발자치코드	0
인허가일자	0	인허가일자	0
영업상태코드	0	영업상태코드	0
상세영업상태	3	상세영업상태	0
소재지	0	소재지	0
도로명	883	도로명	0
사업장명	0	사업장명	0
자본금	0	자본금	0
dtype: int64		dtype: int64	

6. 중복 데이터가 있는 데이터는 첫 번째 데이터는 유지하고 삭제하시오.

※중복 데이터가 성공적으로 삭제되었다면, 다음과 같이 조회되지 않아야 됨 (배점 :5점)

개발자치코드	인허가일자	영업상태코드	상세영업상태	소재지	도로명	사업장명	자본금

7. ‘도로명’ 열을 삭제하시오.

삭제한 후, 결측치가 삭제되었는지 조회해 보세요. (배점 :5점)

개발자치코드	0	개발자치코드	int64
인허가일자	0	인허가일자	object
영업상태코드	0	영업상태코드	int64
상세영업상태	0	상세영업상태	object
소재지	0	소재지	object
도로명	0	사업장명	object
사업장명	0	자본금	int64
자본금	0	dtype: object	
dtype: int64			

8. 열 “자본금” 데이터에 대한 데이터 개수, 평균, 표준편차, 최소값, 25%, 50%, 75, 최대값 통계 값을 출력하시오. (배점 :5점)

```
count      15460.00000
mean       22384826.84347
std        12964949.93065
min         0.00000
25%        11248750.00000
50%        22222000.00000
75%        33550750.00000
max        74500000.00000
Name: 자본금, dtype: float64
```

9. ‘상세영업상태’가 “영업중”을 제외하고 모두 삭제하시오. (배점 :5점)

※조회하였을 ‘상세영업상태’가 “영업중”만 나와야 함

10. 자본금이 30,000,000원 미만인 데이터를 32,000,000원으로 대체하여 적용하시오.

※32,000,000원으로 적용되었으므로 30,000,000 미만인 데이터가 모두 32,000,000으로 대체되어 조회되어야 함 (배점 :5점)

개방자치코드	인허가일자	영업상태코드	상세영업상태	소재지	사업장명	자본금
130	3000000	2021-06-08	1	영업중	서울특별시 종로구 당주동 145	정다운여행사 6723000
131	3000000	2020-11-17	1	영업중	서울특별시 종로구 혜화동 15-19 선릉그린빌리지 101호	둘레길 협동조합 4830000
133	3000000	2021-04-16	1	영업중	서울특별시 종로구 명륜1가 46-3 주호빌딩	다가치포럼 사회적협동조합 10097000
134	3000000	2020-02-27	1	영업중	서울특별시 종로구 수송동 58 두산위브파빌리온	주식회사 웰드미소투어 26360000
135	3000000	2004-03-04	1	영업중	서울특별시 종로구 신문로1가 163번지 광화문오피시아 2105호	(주)투어리아 3251000
...
16277	6520000	2017-07-13	1	영업중	제주특별자치도 서귀포시 안덕면 화순리 318-31 202호	마당밭 투어 18497000
16281	6520000	2016-08-09	1	영업중	제주특별자치도 서귀포시 서홍동 446-5번지	(주)오이오이 3632000
16282	6520000	2010-11-17	1	영업중	제주특별자치도 서귀포시 서홍동 349-8 제1호	투어리스 여행사 10151000
16283	6520000	2001-01-19	1	영업중	제주특별자치도 서귀포시 호근동 1822-5번지	제주도여행사 23764000
16284	6520000	2010-09-09	1	영업중	제주특별자치도 서귀포시 서귀동 298-7	주식회사 한마음투어 25128000

2501 rows × 7 columns



개방자치코드	인허가일자	영업상태코드	상세영업상태	소재지	사업장명	자본금
130	3000000	2021-06-08	1	영업중	서울특별시 종로구 당주동 145	정다운여행사 32000000
131	3000000	2020-11-17	1	영업중	서울특별시 종로구 혜화동 15-19 선릉그린빌리지 101호	둘레길 협동조합 32000000
133	3000000	2021-04-16	1	영업중	서울특별시 종로구 명륜1가 46-3 주호빌딩	다가치포럼 사회적협동조합 32000000
134	3000000	2020-02-27	1	영업중	서울특별시 종로구 수송동 58 두산위브파빌리온	주식회사 웰드미소투어 32000000
135	3000000	2004-03-04	1	영업중	서울특별시 종로구 신문로1가 163번지 광화문오피시아 2105호	(주)투어리아 32000000
...
16277	6520000	2017-07-13	1	영업중	제주특별자치도 서귀포시 안덕면 화순리 318-31 202호	마당밭 투어 32000000
16281	6520000	2016-08-09	1	영업중	제주특별자치도 서귀포시 서홍동 446-5번지	(주)오이오이 32000000
16282	6520000	2010-11-17	1	영업중	제주특별자치도 서귀포시 서홍동 349-8 제1호	투어리스 여행사 32000000
16283	6520000	2001-01-19	1	영업중	제주특별자치도 서귀포시 호근동 1822-5번지	제주도여행사 32000000
16284	6520000	2010-09-09	1	영업중	제주특별자치도 서귀포시 서귀동 298-7	주식회사 한마음투어 32000000

2501 rows × 7 columns

11. ‘인허가일자’에서 연도를 추출하여 ‘개점년도’ 열을 추가하시오. (배점 :10점)

- ① 현재 “인허가일자”는 문자형태의 object 타입인데 날짜형 데이터로 변환하기 위해서는 `pd.to_datetime()` 함수를 적용하면 날짜 형태 자료형인 `datetime` 으로 변환임 가능하다.
- ② 변환된 날짜 데이터에서 년도만 추출하기 위해서는 `df['인허가일자'].dt.year` 명령을 이용한다.(주의! 연도는 `dt.year`이다.)

[Hint]

`df['인허가일자'] = pd.to_datetime(df['인허가일자'])`

`df['개점년도']=df['인허가일자'].dt.year`

```

개방자치코드      int64
인허가일자      datetime64[ns]
영업상태코드      int64
상세영업상태      object
소재지            object
사업장명          object
자본금            int64
dtype: object

```

개방자치코드	인허가일자	영업상태코드	상세영업상태	소재지	사업장명	자본금	개점년도
130	3000000	2021-06-08	1	영업중	서울특별시 종로구 당주동 145	정다운여행사 32000000	2021
131	3000000	2020-11-17	1	영업중	서울특별시 종로구 혜화동 15-19 선릉그린빌리지 101호	둘레길 협동조합 32000000	2020
132	3000000	2020-11-13	1	영업중	서울특별시 종로구 당주동 145	(주)투어람 36418000	2020

12. “년도”를 기준으로 그룹화하여, **대리점 개수**, **년도별 자본금 합계**를 구하시오.
이때 그룹화한 데이터를 새로운 변수에 저장하시오. (배점 :10점)
13. 그룹한 데이터에서 ‘사업장명’을 ‘년도별매장개수’로 이름을 변경하세요. (배점 :5점)

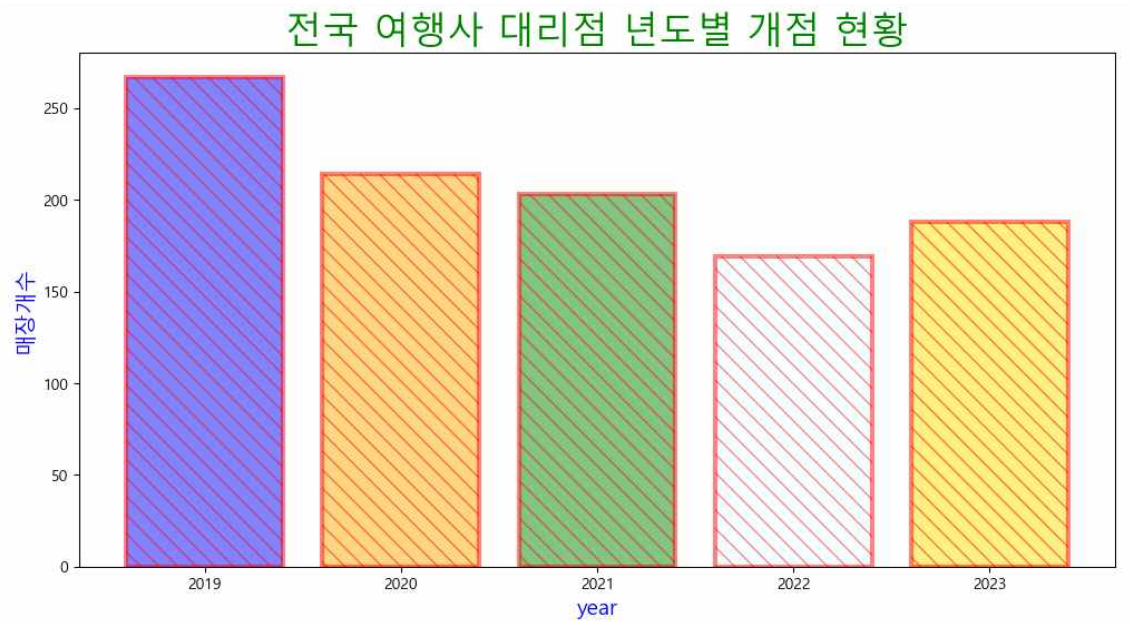
	년도별사업장개수	자본금
개점년도		
1971	1	32000000
1972	1	31282000
1982	1	30566000
1983	3	98045000
1984	2	76018000
1986	4	144877000

14. 그룹한 데이터에서 “년도별사업장개수”가 높은 년도를 기준으로 ‘순위’ 열을 추가하시오.
그 다음 사업장개수가 높은 순으로 출력하시오. (배점 :5점)

	년도별사업장개수	자본금	순위
개점년도			
2019	267	8960112000	1
2018	232	7879218000	2
2015	223	7467657000	3
2020	214	7200691000	4
2021	203	6855779000	5
2017	194	6566727000	6
2023	188	6392128000	7
2016	177	5947425000	8
2022	169	5723777000	9
2013	165	5507895000	10
2014	163	5600762000	11
2012	153	5197824000	12
2011	137	4583368000	13
2007	116	3923317000	14
2010	112	3808906000	15
2008	99	3323830000	16
2009	95	3217359000	17
2024	81	2688727000	18

15. 2018~2023년 사이 “개점년도”, “년도별사업장개수”, “자본금”, “순위” 데이터를 추출하여 새로운 변수 branch_rank에 저장하시오. (배점 :5점)

16. “branch_rank”에 저장된 변수에서, x축은 “개점년도”, y축은 “년도별사업장개수”를 데이터로 활용하여 세로막대 그래프를 표시하는 시각화를 진행하시오. (배점 :10점)



17. 화면에 표시된 그래프를 c:\python_work 폴더에 “travel.png” 파일로 저장하시오. (배점 :5점)

