[ 판다스 실습(6) ]

제출소스명 : lab16.ipynb, hw7.png

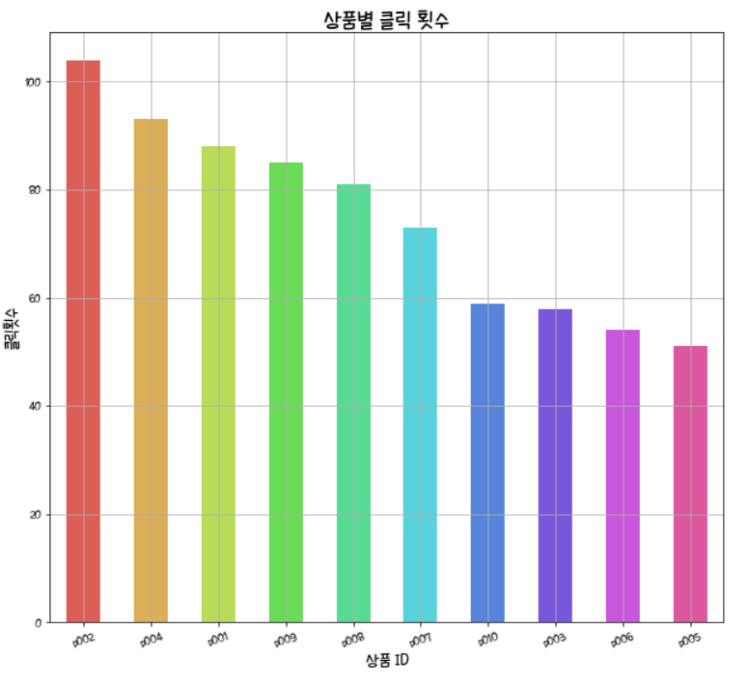
메일제목 : PYTHON-성명-lab16

1. cctv\_seoul.csv 파일 내용을 읽어서 인구수 대비 CCTV수와 인구수 대비 고령자 비율을 계산하여 각각 CCTV비율과 노인비율이라는 열을 추가하고 앞에서 부터 5개 행을 출력한다. CCTV비율과 노인비율 추가시 apply() 함수를 사용해서 추가한다.(교재의 예제 6-5를 참고한다.)



2. product\_click.log 파일을 읽고 데이터프레임으로 생성하여 다음 질문들을 해결해 본다.

(1) 상품별 클릭수를 바그래프로 그리는데 클릭수가 많은 순으로 그린다.(hw7.png에 저장한다)



(2) 어떤 요일에 가장 많이 클릭하는지 다음과 같이 출력하시오.



(3) 어느 시간대에 가장 많이 클릭하는지 다음과 같이 출력하시오.



3. 제공된 emp.csv 를 읽고 데이터프레임으로 생성하여 다음 질문들을 해결해 본다.

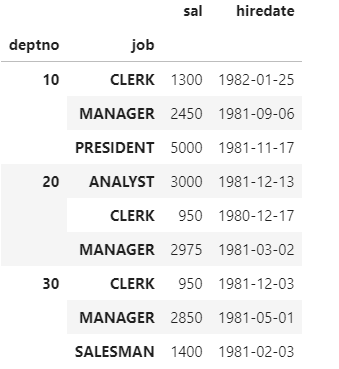
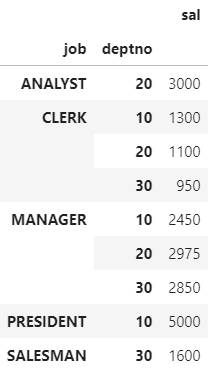
(1) 부서별 월급의 합



(2) 직무(job)별 월급의 합



(3) 부서와 직무(job)별 최고 월급과 입사한지 가장 오래된 직원의 입사날짜



(4) 직무(job)와 부서별 최고 월급

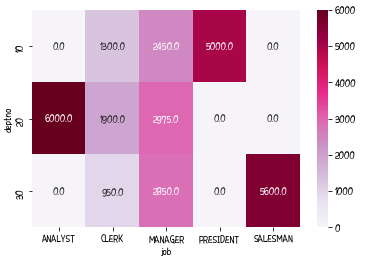
(5) 다음과 같이 부서번호를 인덱스, 직무를 컬럼으로 구성하고 월급의 합으로 구성되는 피벗테이블을 만들고 pv1 변수에 저장한다.



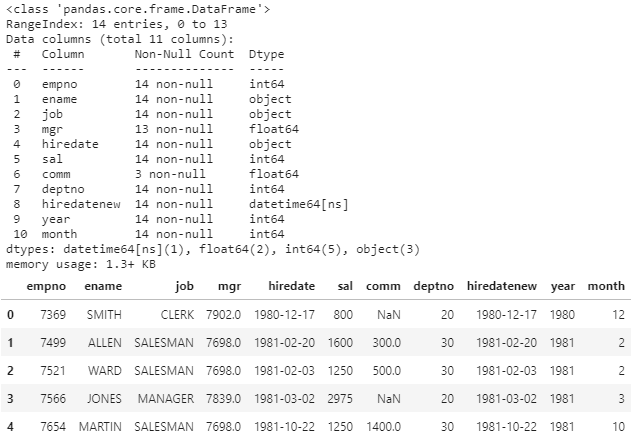
(6) pv1 에서 NaN 값을 0으로 설정한다.



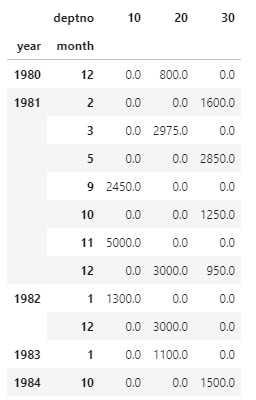
(7) pv1을 가지고 다음과 같은 히트맵을 그린다.



(8) emp 데이터 프레임에서 hiredate 열을 가지고 datetime64 타입의 hiredatenew 열을 만들고 year와 month 열을 추가한 후에 정보를 출력하고 상위 5개 행을 출력한다.

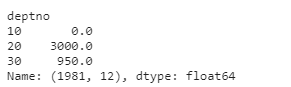


(9) 다음과 같이 입사 년과 월을 인덱스, 직무를 컬럼으로 구성하고 월급의 최대값으로 구성되는 피벗테이블을 만들고 pv2 변수에 저장한다. NaN은 0으로 채운다.

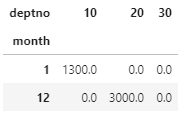


(10) pv2 에서 인덱싱을 통해 다음 데이터들을 추출한다.

- 1981년 12월 데이터들을 추출한다.



- 1982년 데이터들을 추출한다.



- 매년 12월 데이터들을 추출한다.



(11) pv2를 가지고 다음과 같은 히트맵을 그린다.

