**조사방법론2 보고서**

**헌혈 데이터 분석 및 헌혈 장려 어플 개발**

과목명 조사방법론2

교수명 하형태

조 이름 A.S.I (2조)

조원 201832601 김대건

201932133 오승은

202132539 이종관

202132549 한은재

제출일 2022.12.14

목차

**Ⅰ. 서론2**

1. 문제 정의2

2. 주제 설명2

**Ⅱ. 본론3**

**1. 데이터 수집 및 전처리**3

가. 변수 설정3

나. 데이터 웹 크롤링4

**2. 데이터 시각화 및 분석**8

**3. 어플 개발 과정 및 기능**12

가. 개발 목적12

나. 개발 과정12

다. 기능 개발 및 시연13

**Ⅲ. 결론17**

자료 출처17

-

**Ⅰ. 서론**

1. 문제 정의

매년 헌혈 통계에 따르면 우리나라에서는 매년 약 40만명 정도가 외상 및 수술을 비롯한 여러 이유로 수혈을 받고 있다. 그렇기 때문에 의료 시스템 유지에 있어서 헌혈은 매우 필수적이라고 할 수 있다. 혈액관리본부에 따르면, 5일분이 적정 혈액 보유량이기 때문에 장기관 보관이 불가능하다. 또한, 헌혈 조건은 만 17세 이상에서 만 65세 미만의 45kg ~ 50kg 이상인 사람만이 가능하고, 이 외에도 혈액비중, 혈압, 혈소판 수 등의 기준을 통과해야 헌혈이 가능하기 때문에 이러한 이유들로 헌혈 참여가 쉽지 않은 편이다.

특히 2020년 1월부터 지금까지 약 2년 8개월여 지속된 코로나 19로 인해 혈액 보유량의 적정량이 오락가락 하고 있다. 최근 혈액 보유량은 3 ~ 4일분으로 적정 수준인 5일분에 미치지 못하는 ‘관심단계’에 머무르고 있어 혈액 수급에 있어서 어려움을 겪고 있다. 또한, 2016년부터 2020년사이의 5년 동안 적정 혈액 보유량인 날은 514일(28%)에 불과했다. 현재 코로나 19가 많이 완화된 시점이지만, 변이 바이러스로 인한 거짓 소문으로 헌혈 참여율이 더 감소하게 되었다.

2. 주제 설명

프로젝트의 주제는 ‘헌혈 데이터를 분석하고 헌혈 어플을 개발’이다. 앞서 문제 정의 부분을 환기시키면, 여러 가지의 이유로 매년 헌혈률이 감소한다. 헌혈률이 감소한다는 말은, 혈액을 공급받아야 하는 사람은 많아지는데 정작 공급해주는 사람이 줄어든다는 것이다. 특히 우리나라 인구수는 경기도 지역이 약 26%로 가장 많지만, 반대로 전체 헌혈 통계에서 경기도의 비율은 약 8%에 불과한다. 매년 인구수도 줄어들기 때문에 헌혈률이 줄어들고 낮은 수치를 기록할 수밖에 없다. 하지만, 헌혈률의 줄어드는 추세를 막고자 경기도 지역에 한하여 연령, 장소, 직업, 혈액형 등의 헌혈통계 자료 등을 분석하고, 년도에 따라 비교한다. 분석한 결과를 바탕으로, 다양한 기능이 들어간 어플을 개발해 사람들의 헌혈률과 참여를 높이고자 한다.

**Ⅱ. 본론**

**1. 데이터 수집 및 전처리**

가. 변수 설정

테이블이(가) 표시된 사진

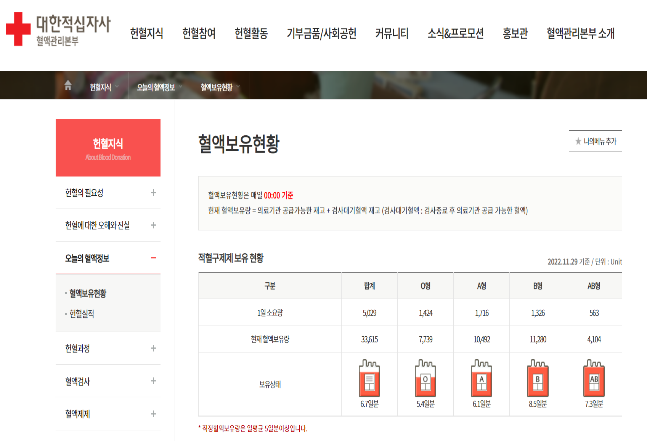
자동 생성된 설명테이블이(가) 표시된 사진

자동 생성된 설명테이블이(가) 표시된 사진

자동 생성된 설명

: 기존 변수는 7개로 월, 직업, 연령, 혈액형, 장소, 헌혈 종류, 성이 있었다. 하지만, 전처리 과정에서 변수 중에 월, 장소, 헌혈 종류가 결측치가 많았고, 다른 변수에 비해 중요하지 않는 변수라고 생각해 최종 변수로 직업, 연령, 혈액형, 성의 4개의 변수로 정했다.

나. 데이터 웹 크롤링 및 전처리



* 대한적십자사 혈액관리본부에서 제공하는 혈액보유현황 데이터를 크롬 개발자 모드와 파이썬을 이용해 웹 크롤링을 했다.

**<웹 크롤링에 이용한 패키지와 함수>**

텍스트이(가) 표시된 사진

자동 생성된 설명

* 웹 크롤링에 이용한 함수 패키지는 requests, BeautifulSoup, urllib, html\_table\_parser, parser 등의 함수 패키지를 불러와 웹 크롤링에 이용했다.

**<url 불러오기, BeautifulSoup 패키지 이용>**

텍스트이(가) 표시된 사진

자동 생성된 설명

* Urllib 패키지를 불러와 urlopen 함수를 이용해 대한적십자사 혈액관리본부의 혈액보유현황 웹 사이트 url을 불러왔다.
* Read 함수를 이용해 텍스트 형태의 데이터를 읽어왔다. 그리고 html 형태로 된 데이터를 추출하기 위해 BeautifulSoup 패키지를 이용했다.

**<table 형태의 html 텍스트 찾는 과정>**

텍스트이(가) 표시된 사진

자동 생성된 설명

* 위에서 BeautifulSoup 패키지를 이용해 html 형태의 텍스트를 추출했는데, 그 중에 표에 있는 적혈구제제 보유현황 테이블 자료가 필요해, html 형태의 텍스트 중 모든 테이블 텍스트를 찾을 수 있는 find\_all(‘table’) 함수를 이용해 테이블 자료를 불러왔다.

**<make2d 기능을 이용해 table 데이터 불러오기>**

텍스트이(가) 표시된 사진

자동 생성된 설명

* Parser\_functions에 있는 make2d를 이용해 table 안에 있는 데이터들을 불러왔다.
* 그리고 pandas, numpy, os의 패키지를 불러오고 set\_index(“구분”)의 함수를 이용해 구분을 정렬해 웹 크롤링으로 불러온 데이터를 table로 나타냈다.

**<df2.csv 파일에 저장하기>**

텍스트이(가) 표시된 사진

자동 생성된 설명

* df2.csv 파일로 저장했다. 이 데이터는 오늘의 혈액정보를 보여주는 앱 기능에 사용된다.

**<연령별 데이터 불러오기, 결측치 제거하기>**

텍스트이(가) 표시된 사진

자동 생성된 설명

* 연령별 데이터를 불러오고 나중에 그래프를 그리는데 이용할 pandas, matplotlib.pyplot, os, matplotlib.font\_manager 패키지를 이용했다.
* Font\_manager 패키지는 그래프에서 폰트가 깨지는 것을 막기 위해 ‘나눔고딕’ 폰트를 이이용했다.
* 결측치를 제거하기 위해 drop 함수를 이용해 모든 열이 결측치인 열을 제거하고, skiprows 함수를 이용해 모든 행이 결측치인 행을 제거했다.

**<년도, 연령, 직업 구분하기>**

텍스트이(가) 표시된 사진

자동 생성된 설명

* Sort 함수를 이용해 list 형태의 데이터를 각각 년도, 연령, 지역으로 정렬해 어떤 데이터가 있는지 확인했다.

**<2019년도와 2021년도 데이터 불러오기>**

텍스트이(가) 표시된 사진

자동 생성된 설명 텍스트이(가) 표시된 사진

자동 생성된 설명

* 위의 불러온 테이블 데이터에서 코로나 이전 상황인 2019년도와 이후 상황인 2021년도를 비교하기 위해 2019년도의 데이터와 2021년도의 데이터를 각각 불러왔다.

**<그래프로 나타내기 위해 년도 열을 제거>**

텍스트이(가) 표시된 사진

자동 생성된 설명 텍스트이(가) 표시된 사진

자동 생성된 설명

* 그래프를 그리려면 년도 열을 삭제해야 해서, drop 함수를 이용해 2019년도와 2021년도의 년도 열을 삭제했다. 그리고 set\_index 함수를 이용해 연령별로 정렬했다.

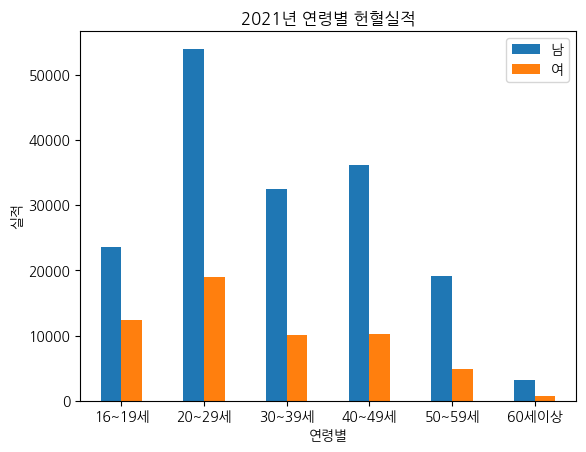
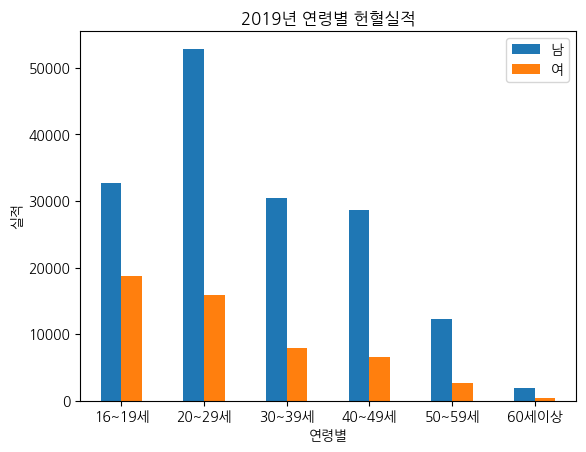
**2. 데이터 분석 및 시각화**

**<2019년과 2021년 시각화한 막대그래프 비교>**

텍스트이(가) 표시된 사진

자동 생성된 설명 텍스트이(가) 표시된 사진

자동 생성된 설명



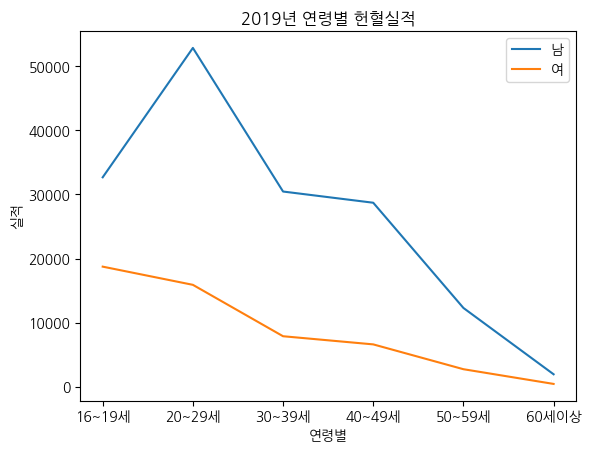
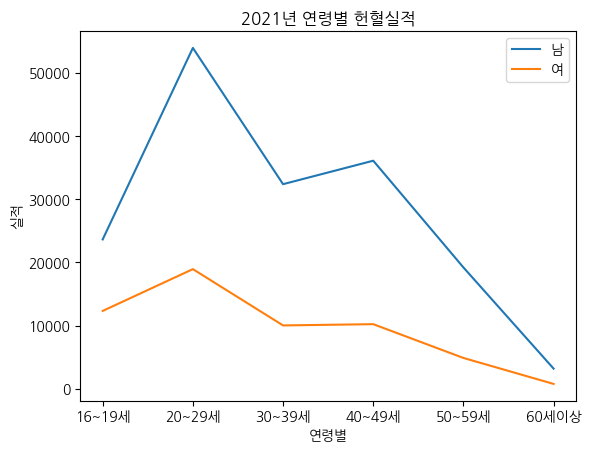
* Groupby 함수를 이용해 각각 2019년도와 2021년도를 그룹으로 묶었다. 그리고 figure(figsize(8,10))를 이용해 그래프의 크기를 정했고, plot(kins=(‘bar’))를 이용해 막대그래프를 그렸다.
* 2019년과 2021년 연령별 데이터를 막대그래프로 시각화 하여 나타냈다. 우리나라의 헌혈을 절반 이상이 10대와 20대가 차지하는데, 16~19세 층과 20대 층에서 헌혈실적이 감소하는 것을 알 수 있다.

**<2019년과 2021년 시각화한 추이선 그래프 비교>**

텍스트이(가) 표시된 사진

자동 생성된 설명텍스트이(가) 표시된 사진

자동 생성된 설명

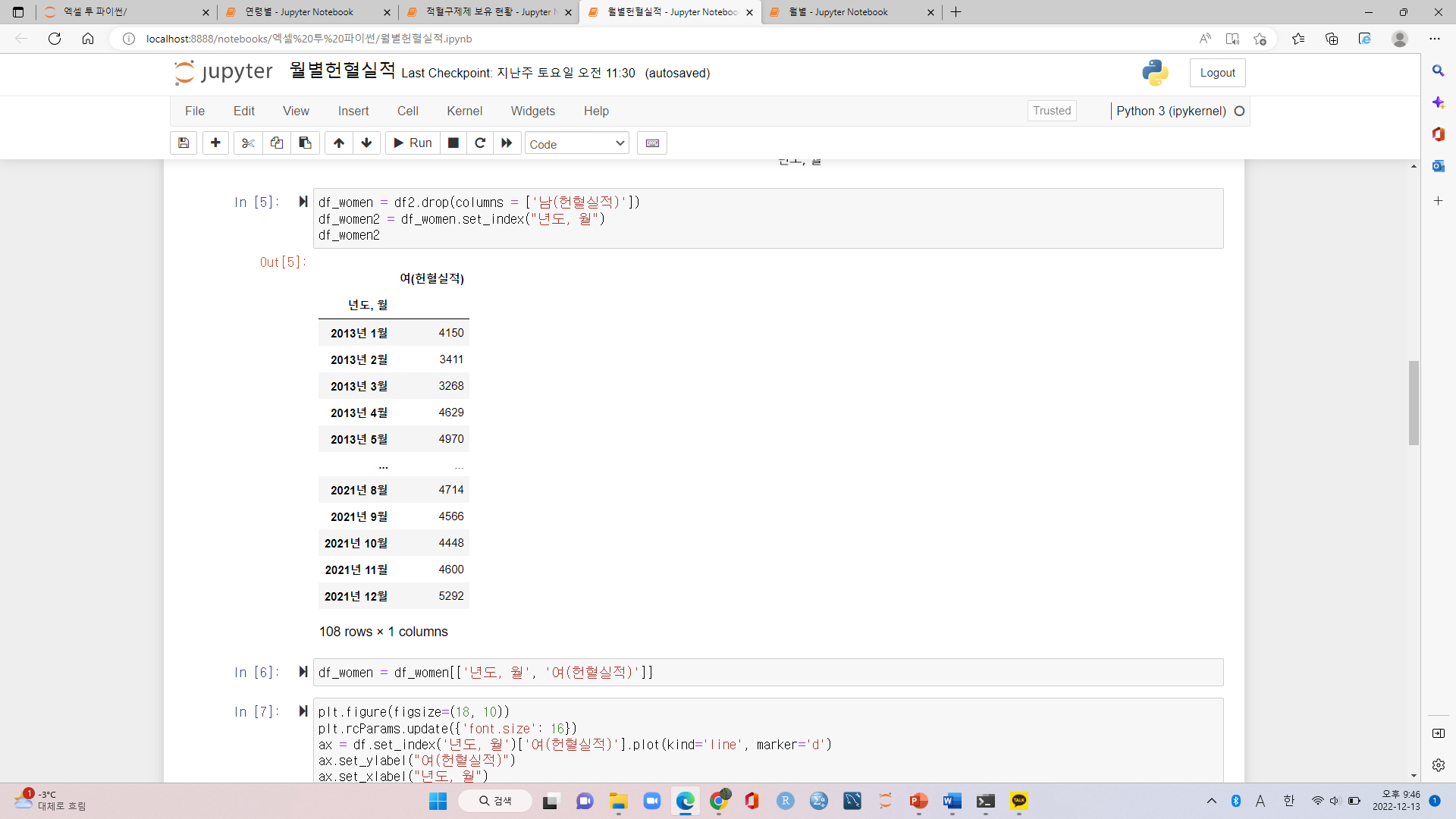
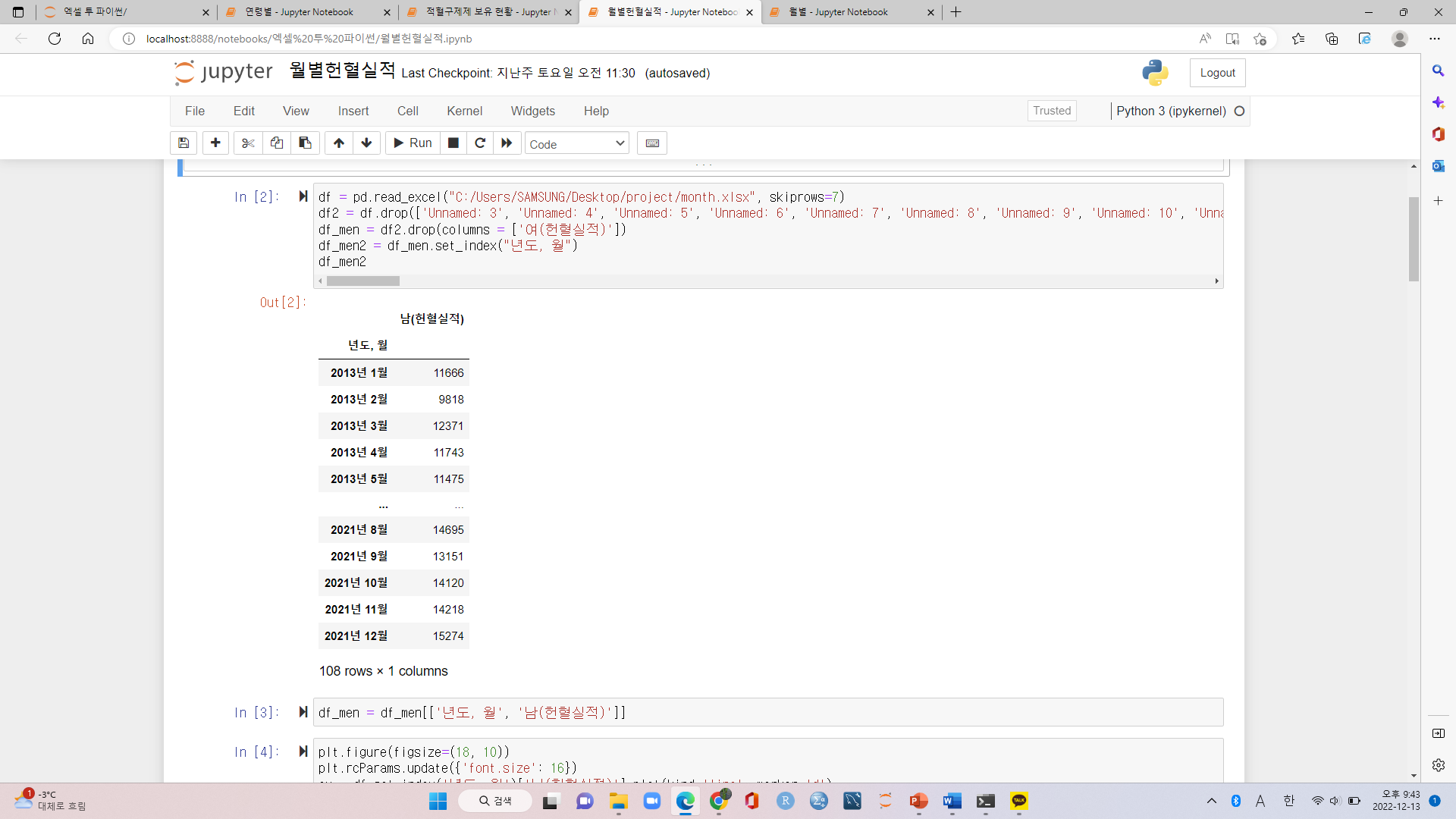
**<성별에 따른 헌혈실적 비교>**

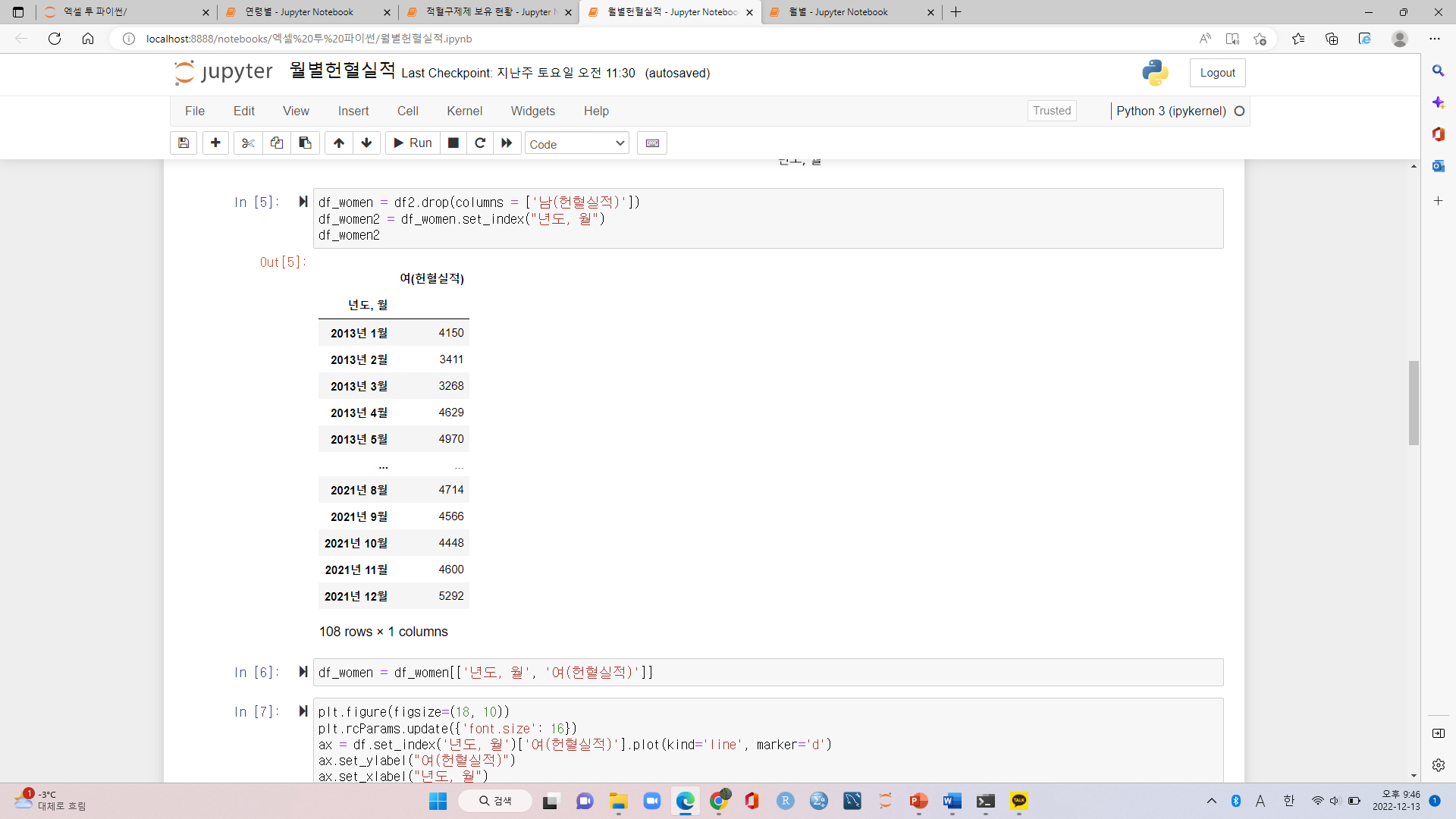
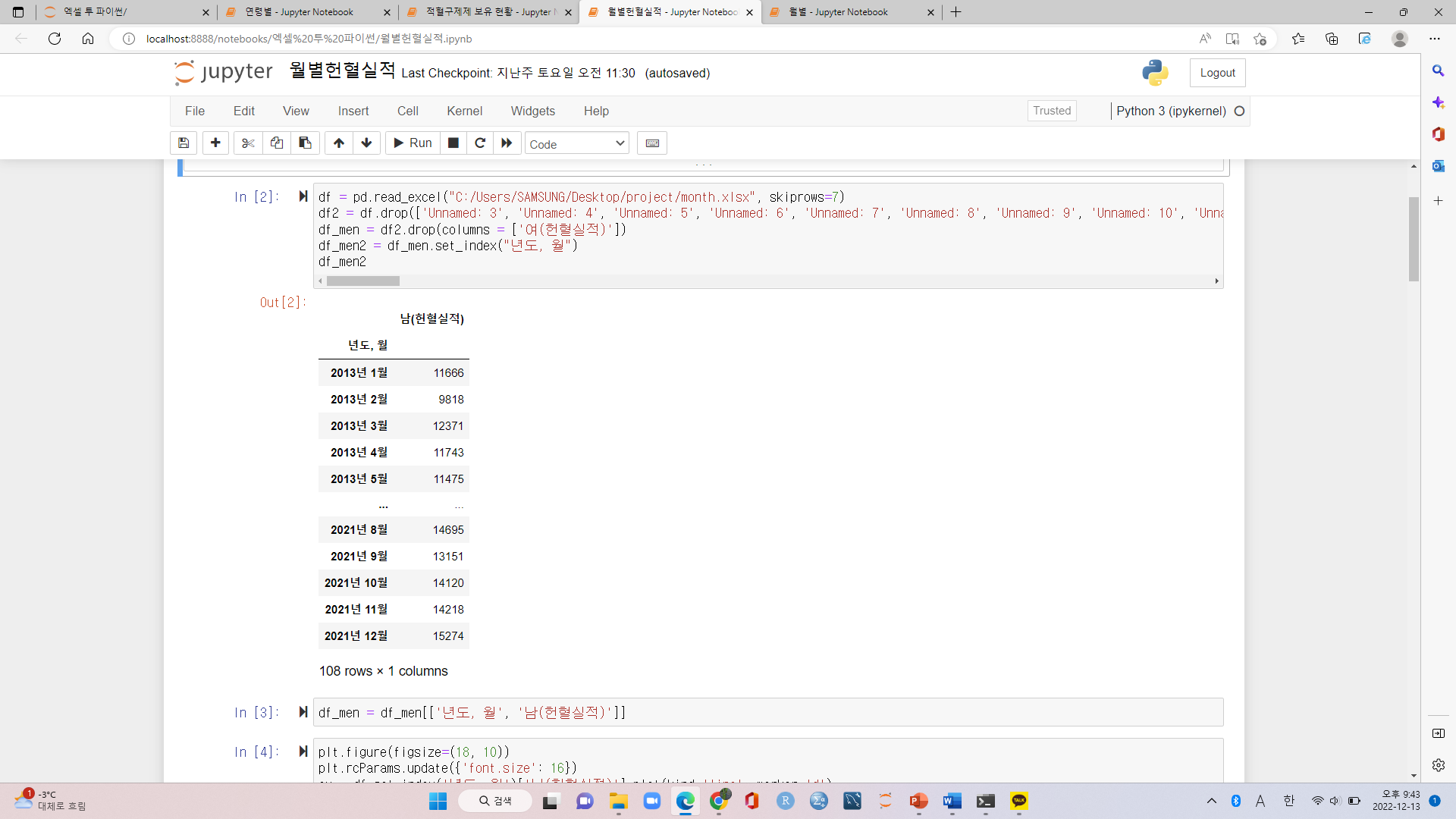
텍스트이(가) 표시된 사진

자동 생성된 설명

* 성별에 따른 월별 추이선 그래프를 그리기 위해 pandas, matplotlib.pyplot, os, matplotlib.font\_manager 패키지를 이용했다.

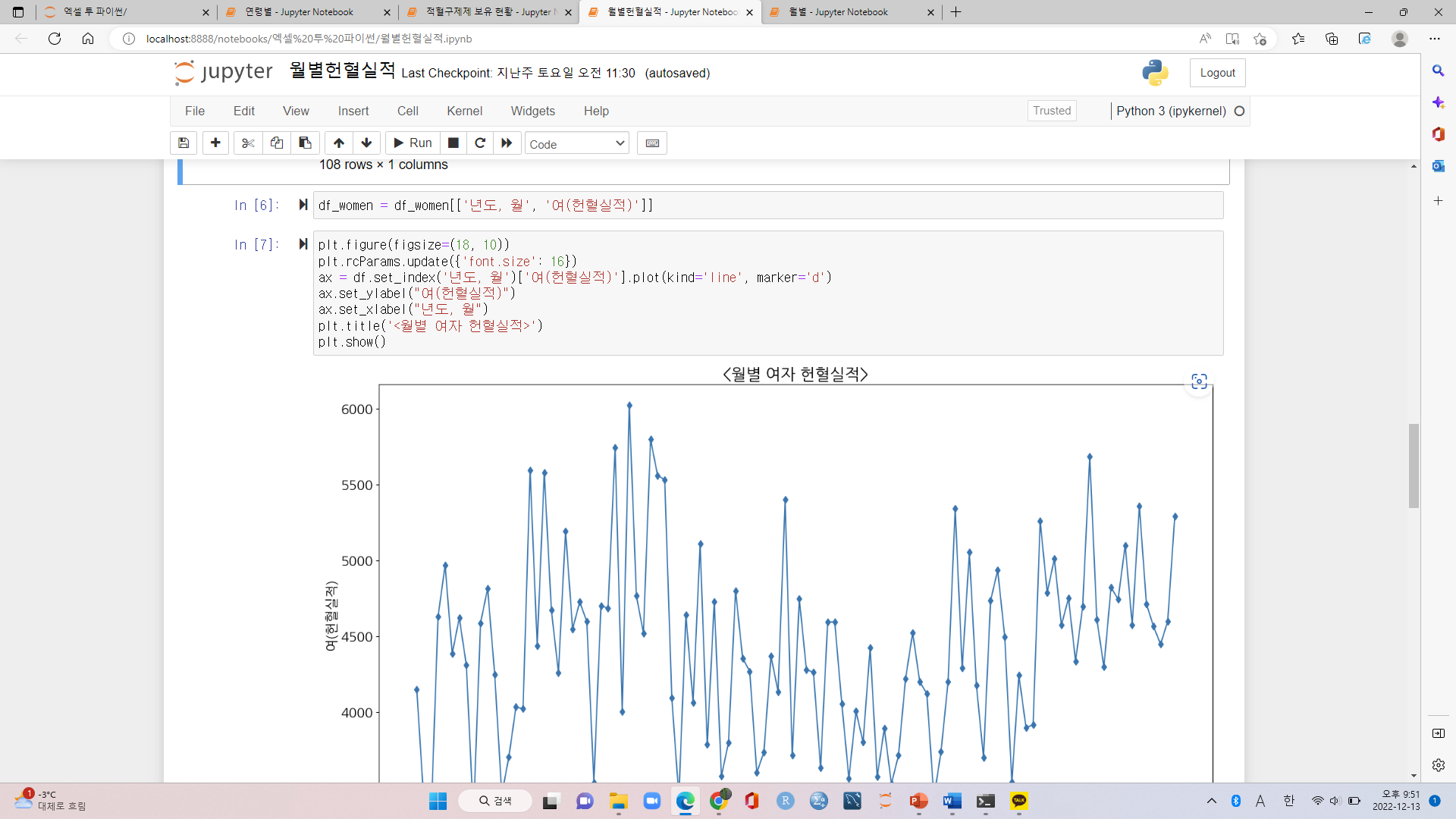
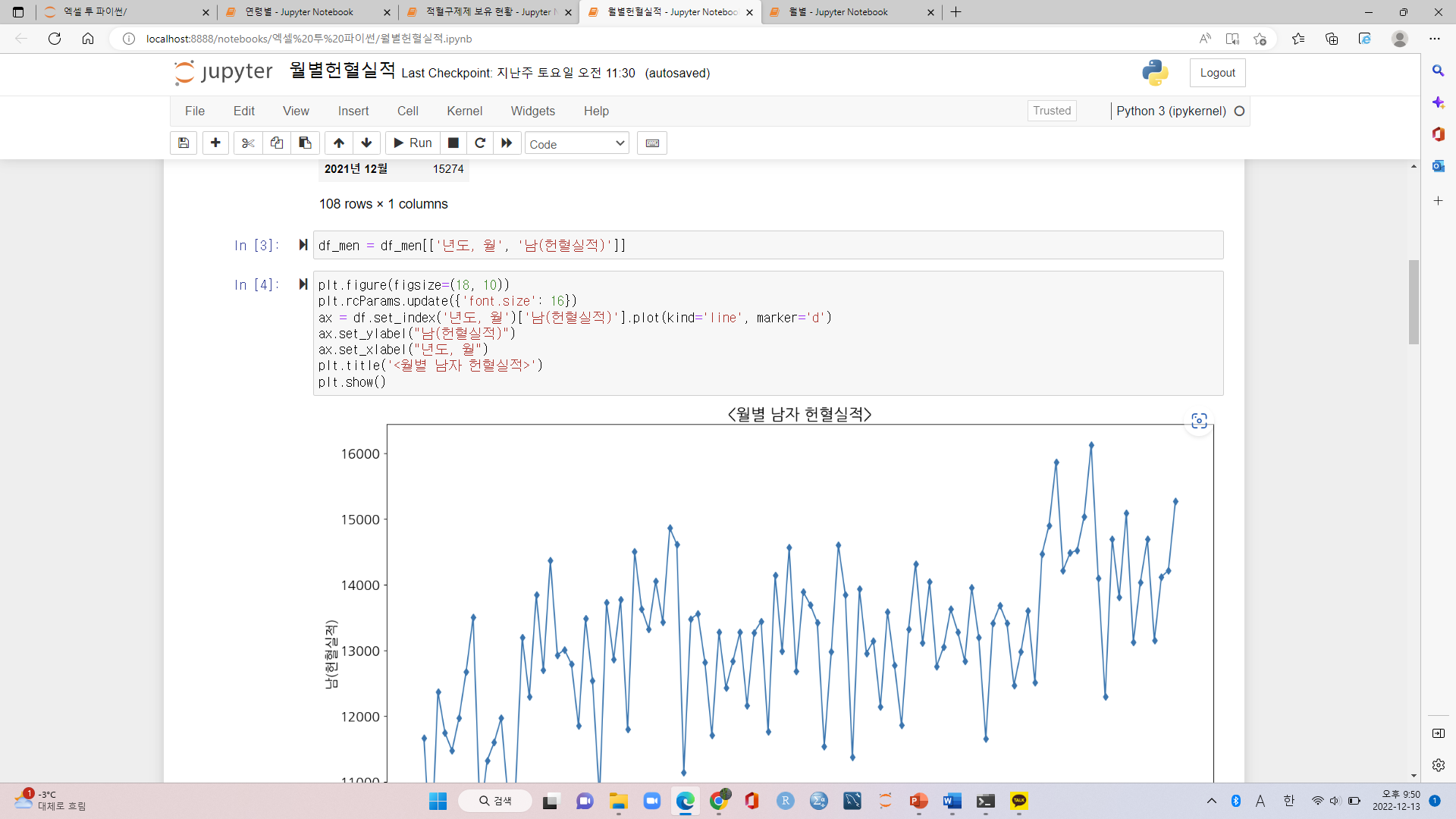
**<성별에 따른 2013년~2021년의 월별 헌혈실적 데이터 불러오기, 전처리 과정>**

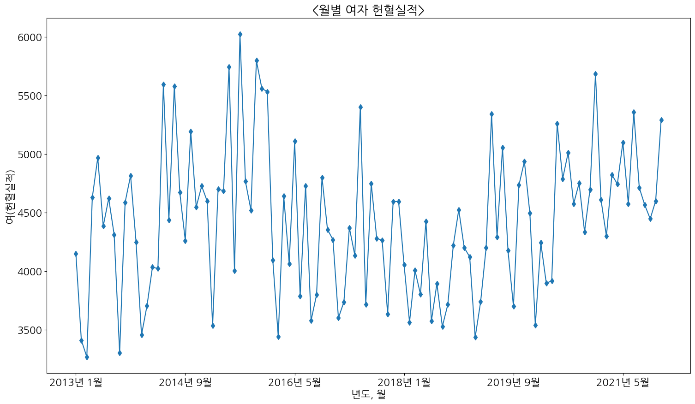
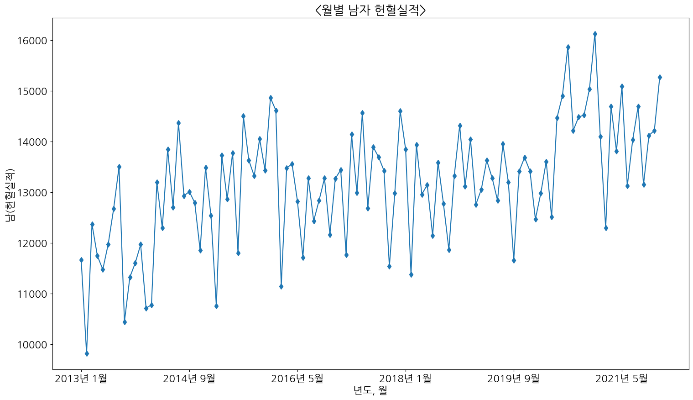




* Drop 함수로 모든 열이 결측치인 열을 제거하고, skiprows 함수를 이용해 모든 행이 결측치인 행을 제거했다.
* Set.index 함수를 이용해 년도, 월로 정렬했다.

**<가로축: 년도, 월, 세로축: 남녀 헌혈실적으로 정렬한 후 추이선 그래프 그리기>**





* Set.index 함수를 이용해 가로축을 년도, 월, 세로축을 남녀의 헌혈실적으로 했고, plot(kind=’line’, marker=’d’)를 이용해 추이선 그래프에 마커를 점으로 표시했다.
* 양의 상관관계로 그래프의 형태가 비슷한 것을 확인할 수 있다.

**<탐색적자료분석(EDA)을 이용한 상관계수 구하기>**

텍스트이(가) 표시된 사진

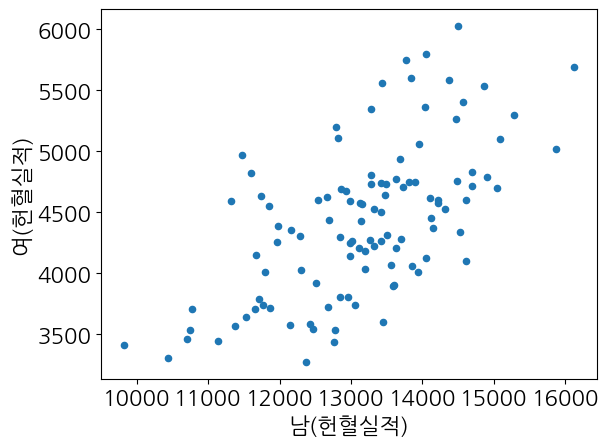
자동 생성된 설명텍스트이(가) 표시된 사진

자동 생성된 설명

* Describe 함수로 탐색적자료분석(EDA)으로 평균, 표준편차, 최소 최대값, 사분위수 값을 구했다. 그리고 corr 함수로 남자의 헌혈실적과 여자의 헌혈실적을 구했다.
* 남자의 헌혈실적과 여자의 헌혈실적의 상관계수는 0.61516으로 강한 양의 상관관계이다.

**<성별에 따른 헌혈실적을 산점도로 나타내기>**





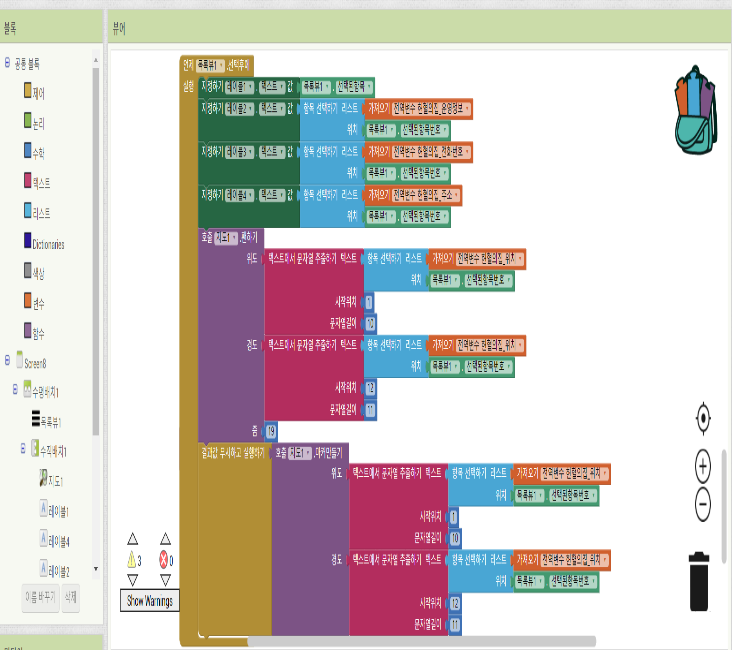
* Plot(kind=’scatter’, x=’남(헌혈실적)’, y=’여(헌혈실적)’)의 함수를 이용해 x축은 남자의 헌혈실적, y축은 여자의 헌혈실적으로 설정해 산점도를 구했다. 위의 상관계수를 구한 것처럼 양의 상관관계인 것을 확인할 수 있다.

**3. 어플 개발 과정 및 기능**

가. 개발 목적

전국에서 경기도의 헌혈률이 가장 낮은 점을 고려하여, 경기도민들이 쉽고 편하게 헌혈을 할 수 있도록 돕는 어플을 개발하였다. 사용자에게 헌혈 현황을 알려주어 헌혈의 동기를 제공하고, 위치정보를 이용하여 가장 가까운 헌혈의 집까지의 자세한 경로를 알 수 있도록 하여 헌혈 참여를 끌어올릴 수 있도록 하였다.

나. 개발 과정

노코딩툴인 ‘MIT App Inventor’(이하 ‘앱 인벤터’라 한다)를 이용하여 개발하였다. 앱 인벤터는 코딩 없이 Android기반 어플을 무료로 만들 수 있는 프로그램으로, 웹사이트를 이용하여 손쉽게 어플을 개발할 수 있도록 돕는다. 원하는 기능을 보여주는 화면을 디자이너 탭에서 구현하고, 블록 탭에서 여러 기능을 가지고 있는 블록을 조립, 배치하여 직관적으로 구현할 수 있다. 블록에는 제어, 논리, 수학, 텍스트 등의 종류가 있으며, 카메라, 녹음, 위치 및 방향 센서 등을 이용할 수 있다.

↳좌 – 디자이너 탭 , 우- 블록 탭

다. 기능 개발 및 시연

4개의 페이지로 구성되어 있으며, 1)오늘의 헌혈 현황, 2) 혈액보유현황 추이 페이지들은 매일 00시 기준으로 적혈구 제제와 농축 혈소판 보유량 및 최근 1달 간의 추이 그래프를 보여준다. 3) 헌혈의 집 경로 검색, 4) 헌혈의 집 정보 검색 페이지들은 사용자의 위치 정보를 기반으로 가장 가까운 헌혈의 집과 경로 및 전화번호 등의 상세 정보를 표시하는 기능을 가진다.

텍스트이(가) 표시된 사진

자동 생성된 설명텍스트이(가) 표시된 사진

자동 생성된 설명텍스트, 도구, 스크린샷이(가) 표시된 사진

자동 생성된 설명1) 오늘의 헌혈 현황 (1페이지)

↳좌측부터 [그림1], [그림2], [그림3]

[그림1]은 어플을 구동하였을 때 나타나는 화면으로, 버튼을 클릭하여 보유 현황을 볼 수 있다. [그림2], [그림3]은 각각 적혈구 및 농축혈소판 보유 현황 버튼을 눌렀을 시 결과물로, 일일 혈액 보유 현황을 텍스트 형태로 나타낸다. 앞서 데이터 부분에서 혈액관리본부 웹사이트를 Python의 BeautifulSoup, requests 패키지를 이용하여 웹 크롤링하였고, 혈액정보를 csv파일 형태로 추출하였다. 이 데이터 파일을 다시 앱 인벤터에 업로드하고, csv파일을 리스트 형태로 출력하도록 하였다. 휴대폰의 뒤로 가기 버튼을 누르면 다시 처음 화면을 돌아온다.

텍스트이(가) 표시된 사진

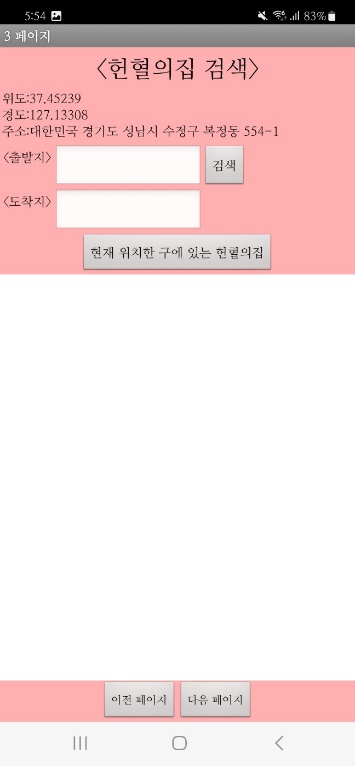
자동 생성된 설명텍스트이(가) 표시된 사진

자동 생성된 설명2) 혈액보유현황 추이 (2페이지)

↳좌측부터 [그림4], [그림5], [그림6]

1페이지에서 다음페이지 버튼을 누르면 [그림4]의 화면이 나타난다. 위의 적혈구제제 보유 추이 버튼을 누르면 [그림5]와 같이 1달 간의 적혈구제제 보유 추이 그래프가 화면 중앙에 나타나며, 농축적혈구 보유 추이 버튼을 누르면 [그림6]과 같이 1달 간의 농축적혈구 보유 현황 추이 그래프가 화면 중앙에 나타난다. 위의 그래프는 앱 인벤터에 이미지 파일을 업로드 하여 구현한 것으로, 혈액관리본부 웹 크롤링 자료를 정리하고 파이썬으로 시각화한 것을 캡쳐하여 이미지 파일로 이용하였다.

텍스트이(가) 표시된 사진

자동 생성된 설명3) 헌혈의 집 경로 검색(3페이지)

↳좌측부터 [그림7], [그림8]

지도이(가) 표시된 사진

자동 생성된 설명텍스트이(가) 표시된 사진

자동 생성된 설명2페이지에서 다음 페이지 버튼을 누르면 [그림7]의 화면이 나타난다. 사용자가 위치정보 이용에 동의하면 위와 같이 위도, 경도 및 주소가 출력되며, ‘현재 위치한 구에 있는 헌혈의 집’ 버튼을 누르면 [그림8]과 같이 주소에 기반하여 근처의 헌혈의 집 리스트를 출력한다. 사용자는 리스트를 통해 가장 가까운 헌혈의 집의 위치를 알 수 있다. [그림8]은 성남시 수정구 가천대학교를 중심으로 거리를 산출한 것으로, 헌혈의 집 매봉 센터가 11km 떨어져 있어 가장 가까운 곳임을 알 수 있다.

↳좌측부터 [그림9], [그림10]

가장 가까운 곳이 매봉센터임으로, 가천대학교를 출발지로 입력하고 도착지에 매봉센터를 입력하면 [그림9]와 같이 최단시간 경로를 안내한다. 경로를 클릭하면 [그림10]과 같이 지도 상의 경로로 볼 수도 있다. 현재 위치한 구에 있는 헌혈의 집은 네이버지도를 통해 산출하고, 경로 찾기는 구글맵을 통해 산출하여 웹뷰어를 통해 보여주는 방식으로 기능을 구현하였다.

지도이(가) 표시된 사진

자동 생성된 설명지도이(가) 표시된 사진

자동 생성된 설명4) 헌혈의 집 정보 검색

↳좌측부터 [그림11], [그림12]

3페이지에서 다음 페이지 버튼을 누르면 [그림11]의 화면이 나타난다. 좌측 리스트에서 경기도 내 13개의 헌혈의 집 목록을 스크롤하여 확인할 수 있으며, 하나를 고르면 [그림12]와 같이 지도 상의 위치가 마커와 함께 표시된다, 지도 아래에는 주소, 운영시간, 전화번호 및 운영일 등의 상세정보를 보여주어, 사용자가 헌혈의 집에 대한 정보를 쉽고 빠르게 얻을 수 있다. 공공데이터 포털에서 전국 헌혈의 집 위치 데이터를 받아 경기도에 위치한 헌혈의 집들을 뽑아낸 후, 앱 인벤터에 텍스트 형태로 입력하였다. 지도 기능을 통해 위치를 마커로 표시하고, 아래에 텍스트 레이블을 두어 헌혈의 집 정보를 알 수 있도록 기능을 구현하였다.

**<결론>**

: 남자와 여자의 헌혈실적의 상관계수를 구해본 결과 강한 양의 상관관계가 나타났다. 두 성별 중에 하나의 성별의 헌혈실적이 증가하면 다른 성별도 증가하고, 다른 하나의 성별의 헌혈실적이 감소하면 다른 성별도 감소하는 관계인 정비례 관계를 알 수 있었다. 그리고 월별, 연령별, 혈액형별, 장소별, 직업별 데이터로 코로나 전후의 상황인 2019년과 2021년을 막대그래프와 추이도로 비교하면서, 코로나 이전 상황보다 이후 상황에서 헌혈실적이 감소한 것을 확인할 수 있었다.

오늘의 헌혈 현황과 혈액 보유 현황 추이, 헌혈의집 검색, 헌혈의집 리스트 기능으로 경기도 지역 사람들의 헌혈 참여를 권장하도록 한다. 사람들에게 앱을 통해, 오늘의 헌혈 현황과 혈액 보유 현황 추이 기능으로 부족한 헌혈참여와 혈액량의 부족을 알리고, 헌혈의집 검색과 리스트 기능을 통해 근처의 헌혈의집을 검색하거나 경기도 지역의 헌혈의집을 검색하여 헌혈 참여 증가시킬 수 있을 것이다.

앞으로 시간이 지날수록 혈액 공급과 헌혈자는 줄어들고 혈액의 수요는 늘어날 것이다. 나 자신도 미래에 혈액이 부족해 다른 사람에게 도움을 받아야 하는 상황이 올 지도 모른다. 헌혈률이 계속 떨어지는 것을 막기 위해 이 앱을 통해 조금이나마 헌혈률이 상승하기를 기대해본다.

**<출처>**

* 국가통계포털 KOSIS(월별, 연령별, 혈액형별, 장소별, 직업별 헌혈통계)
* 대한적십자사 혈액관리본부(적혈구제제 보유 현황, 농축혈소판 보유 현황, 적혈구제제 보유 현황 추이, 농축혈소판 보유 현황 추이)
* 연간 혈액 보유일수 현황<https://www.akomnews.com/bbs/board.php?bo_table=news&wr_id=46446> 대한적십자사 제출 자료 '2016~2020년 혈액보유일수 현황'
* 헌혈 가능 인구와 실 인원수 추이 - KOSIS 혈액정보통계 - 헌혈자수(2005~2021)
* 수혈용 혈액 자급률 예측(2019-2030) - '혈액 및 제대혈 관리실태' 감사원 감사보고서
* 경기도의 인구수 비중, 전체 헌혈 통계에서 경기도의 비중 - KOSIS 혈액정보통계 - 시, 도별 인구 대비 헌혈실적