SAMSUNG SW ACADEMY FOR YOUTH

함께가요 미래로!
Enabling People

삼성 청년 SW 아카데미

데이터분석 이론 및 실습

Confidential

<알림>

본 강의는 삼성 청년 SW아카데미의 컨텐츠로
보안서약서에 의거하여
강의 내용을 어떠한 사유로도 임의로 복사, 촬영,
녹음, 복제, 보관, 전송하거나
허가 받지 않은 저장매체를
이용한 보관, 제3자에게 누설, 공개,
또는 사용하는 등의 행위를 금합니다.

SAMSUNG SW ACADEMY FOR YOUTH

함께가요 미래로!
Enabling People

9월11일 실습 및 과제 알고리즘

SAMSUNG SW ACADEMY FOR YOUTH

함께가요 미래로!
Enabling People

실습 알고리즘

실습1. 타이타닉 데이터셋을 이용한 기본적인 데이터 조작 및 탐색

• 1-1. Titanic 데이터셋 로드 및 구조 확인

1st. Titanic 데이터셋을 로드하기 위해 pd.read_csv() 함수를 사용합니다

2nd. 데이터를 로드한 후, head() 메서드를 사용하여 첫 5행을 출력하여 데이터의 구조를 확인합니다

• 1-2. 특정 열 선택하여 새로운 데이터프레임 생성

1st. 'Name', 'Sex', 'Age', 'Pclass', 'Fare' 열만 선택하기 위해 열 이름을 리스트로 지정합니다.

2nd. 해당 열들을 기존 데이터프레임에서 추출하여 새로운 데이터프레임을 만듭니다.

• 1-3. Age열의 결측치 확인

1st. Age 열에서 결측치가 있는지 확인하기 위해 isnull().sum() 메서드를 사용합니다.

2nd. 결측치가 있다면 이를 출력하여 데이터의 완전성을 파악합니다.

• 1-4. Pclass 별 승객 수 계산

1st. 각 클래스에 몇 명의 승객이 탑승했는지 확인하기 위해 value_counts() 메서드를 사용합니다

2nd. Pclass 열의 값을 기반으로 각 등급별 승객 수를 계산합니다

- 1-5. 요금이 0보다 큰 승객 필터링

1st. 각 Fare 열에서 요금이 0보다 큰 승객들만 선택하기 위해 조건문을 사용합니다

2nd. 조건에 맞는 승객들을 새로운 데이터프레임으로 만들어, 요금이 실제로 지불된 데이터를 확보합니다

실습2. 타이타닉 데이터셋에서 열 및 행 선택 방법 연습

• 2-1. 특정 열 선택하여 새로운 데이터프레임 생성

1st. Titanic 데이터셋에서 'Name', 'Age', 'Sex' 열만 선택하기 위해 열 이름을 리스트로 지정합니다

2nd. 지정된 열들을 기존 데이터프레임에서 추출하여 새로운 데이터프레임을 생성합니다

실습2. 타이타닉 데이터셋에서 열 및 행 선택 방법 연습

• 2-2. 30살 이상인 승객 필터링

1st. Age가 30살 이상인 승객만 선택하기 위해 titanic_data['Age'] >= 30 조건문을 사용합니다

2nd. 조건에 맞는 데이터를 필터링하여 새로운 데이터프레임을 생성합니다

실습2. 타이타닉 데이터셋에서 열 및 행 선택 방법 연습

- ## 2-3. 3등급 중 요금이 20 이하인 승객 필터링

1st. Pclass가 3이고 Fare가 20 이하인 승객을 선택하기 위해 (titanic_data['Pclass'] == 3) & (titanic_data['Fare'] <= 20) 조건문을 사용합니다

2nd. 조건에 맞는 승객들을 필터링하여 새로운 데이터프레임을 생성합니다

실습2. 타이타닉 데이터셋에서 열 및 행 선택 방법 연습

- ## 2-4. 40살 이상이고 1등급인 승객 필터링

1st. Age가 40살 이상이고 Pclass가 1인 승객을 선택하기 위해 (titanic_data['Age'] >= 40) & (titanic_data['Pclass'] == 1) 조건문을 사용합니다

2nd. 조건에 맞는 승객들의 'Name', 'Age', 'Pclass' 열만 선택하여 새로운 데이터프레임을 생성합니다

실습3. 타이타닉 데이터셋에서 조건에 따른 데이터 필터링

• 3-1. 생존한 승객 필터링

1st. 생존한 승객만 필터링하기 위해 Survived 열이 1인 조건을 사용합니다

2nd. 조건에 맞는 데이터를 추출하여 생존자 데이터프레임을 만듭니다

실습3. 타이타닉 데이터셋에서 조건에 따른 데이터 필터링

• 3-2. 여성 및 18세 이하 승객 필터링

1st. 여성 승객을 선택하기 위해 Sex 열이 'female'인 조건을 사용합니다

2nd. 18세 이하 승객을 선택하기 위해 Age 열이 18 이하인 조건을 추가하여, 두 조건을 결합해 데이터를 필터링합니다

실습3. 타이타닉 데이터셋에서 조건에 따른 데이터 필터링

• 3-3. Fare 상위 10% 승객 필터링

1st. 상위 10%에 속하는 Fare의 기준값을 계산하기 위해 quantile(0.9) 메서드를 사용합니다

2nd. 계산된 기준값을 초과하는 Fare 값을 가진 승객들을 필터링하여 상위 요금 지불자 그룹을 추출합니다

3rd. 필터링된 승객 수를 출력하여 상위 10%에 속하는 승객 수를 확인합니다

• 4-1. Age, Fare 열의 기본 통계량 계산

1st. Age와 Fare 열의 평균을 계산하기 위해 mean() 함수를 사용합니다

2nd. Age와 Fare 열의 중앙값을 계산하기 위해 median() 함수를 사용합니다

3rd. Age와 Fare 열의 최솟값을 계산하기 위해 min() 함수를 사용합니다

4th. Age와 Fare 열의 최댓값을 계산하기 위해 max() 함수를 사용합니다

• 4-2. Pclass별 평균 Fare 계산 및 정렬

1st. Pclass별 평균 Fare를 계산하기 위해 groupby('Pclass')와 mean() 함수를 사용합니다

2nd. 계산된 평균 Fare 값을 내림차순으로 정렬하기 위해 sort_values(ascending=False)를 사용합니다

• 4-3. 생존 여부에 따른 Age의 평균 계산

1st. 생존 여부에 따라 Age의 평균을 계산하기 위해 groupby('Survived')와 mean() 함수를 사용합니다

2nd. 생존자와 비생존자의 나이 평균을 비교하여 생존에 나이가 미친 영향을 분석합니다

실습5. 타이타닉 데이터의 정렬 및 상위 데이터 추출

• 5-1. Fare를 기준으로 내림차순 정렬 후 상위 5명 출력

1st. Fare를 기준으로 데이터를 내림차순 정렬하기 위해 sort_values(by='Fare', ascending=False)를 사용합니다

2nd. 정렬된 데이터에서 상위 5명의 승객 정보를 추출하기 위해 head(5)를 사용합니다

실습5. 타이타닉 데이터의 정렬 및 상위 데이터 추출

- ## 5-2. Age가 어린 순으로 정렬 후 상위 10명 출력

1st. Age를 기준으로 오름차순 정렬하기 위해 sort_values(by='Age', ascending=True)를 사용합니다

2nd. 정렬된 데이터에서 나이가 어린 상위 10명의 승객 정보를 추출하기 위해 head(10)을 사용합니다

실습5. 타이타닉 데이터의 정렬 및 상위 데이터 추출

- 5-3. Pclass와 Fare 기준으로 오름차순 정렬 후 상위 10명 출력

1st. Pclass와 Fare를 동시에 오름차순으로 정렬하기 위해 sort_values(by=['Pclass', 'Fare'], ascending=[True, True])를 사용합니다

2nd. 정렬된 데이터에서 상위 10명의 승객 정보를 추출하기 위해 head(10)을 사용합니다

실습5. 타이타닉 데이터의 정렬 및 상위 데이터 추출

- 5-4. 생존자 중 Fare가 높은 순으로 정렬 후 상위 5명 출력

1st. 생존자만 선택하기 위해 Survived가 1인 승객들을 필터링합니다

2nd. 필터링된 생존자 데이터를 Fare 기준으로 내림차순 정렬한 후, 상위 5명을 추출합니다

실습5. 타이타닉 데이터의 정렬 및 상위 데이터 추출

- 5-5. Age, Fare, Pclass 기준으로 가중 평균 계산 후 상위 5명 출력

1st. Age, Fare, Pclass의 가중 평균을 계산하기 위해 각각 0.3, 0.5, 0.2의 가중치를 적용하여 새로운 열을 생성합니다

2nd. 가중 평균 점수를 기준으로 내림차순 정렬하여, 상위 5명의 승객을 추출합니다. 이들이 우선 구조될 대상자들입니다

SAMSUNG SW ACADEMY FOR YOUTH

함께가요 미래로!
Enabling People

과제 알고리즘

과제1. 데이터 프레임 생성 및 기본적인 조작

• 1-1. 특정 열을 선택하여 새로운 데이터프레임 생성

1st. Titanic 데이터셋에서 'Name', 'Sex', 'Age', 'Pclass', 'Fare' 열만 선택하기 위해 열 이름을 리스트로 지정합니다

2nd. 지정된 열을 기존 데이터프레임에서 추출하여 새로운 데이터프레임을 생성합니다

과제1. 데이터 프레임 생성 및 기본적인 조작

• 1-2. 나이에 따라 분류하는 열 추가

1st. 나이에 따라 승객을 'Child', 'Adult', 'Senior'로 분류하기 위해 'AgeGroup'이라는 새로운 열을 추가합니다

2nd. 나이가 0-18세인 승객은 'Child', 19-60세는 'Adult', 61세 이상은 'Senior'로 분류합니다

과제1. 데이터 프레임 생성 및 기본적인 조작

• 1-3. 요금이 높은 승객 추출

1st. 요금이 50 이상인 승객들을 선택하기 위해 'Fare' 열에서 조건을 설정합니다

2nd. 조건에 맞는 승객들을 필터링하여 새로운 데이터프레임을 생성합니다

• 1-4. Pclass가 3인 승객의 평균 나이 계산

1st. Pclass가 3인 승객들을 선택하기 위해 Pclass 열이 3인 조건을 설정합니다

2nd. 조건에 맞는 승객들의 Age 평균을 계산하여 결과를 출력합니다

• 1-5. 여성 승객의 평균 Fare 계산

1st. 여성 승객을 선택하기 위해 Sex 열이 'female'인 조건을 설정합니다

2nd. 조건에 맞는 승객들의 평균 Fare를 계산하여 결과를 출력합니다

과제2. 조건에 따른 데이터 필터링 및 기본 통계량 계산

• 2-1. 생존자 필터링 후 데이터프레임 생성

1st. 생존한 승객들만 필터링하기 위해 Survived 열이 1인 조건을 설정합니다

2nd. 조건에 맞는 승객들을 새로운 데이터프레임으로 생성하여 생존자 그룹을 분석할 수 있게 합니다

• 2-2. Pclass별 생존자 평균 Fare 계산

1st. 생존자 데이터프레임을 Pclass로 그룹화하기 위해 groupby('Pclass')를 사용합니다

2nd. 각 그룹의 평균 Fare를 계산하여 결과를 출력합니다

• 2-3. 30세 이상 승객 필터링

1st. 나이가 30세 이상인 승객들을 선택하기 위해 Age 열이 30 이상인 조건을 설정합니다

2nd. 조건에 맞는 승객들을 필터링하여 새로운 데이터프레임을 생성합니다

과제2. 조건에 따른 데이터 필터링 및 기본 통계량 계산

• 2-4. 특정 조건의 Age 통계량 계산

1st. Pclass가 1이고, Sex가 'male'인 승객들을 필터링하기 위해 두 조건을 결합합니다

2nd. 필터링된 승객들의 평균과 중앙값을 계산하여 결과를 출력합니다

과제2. 조건에 따른 데이터 필터링 및 기본 통계량 계산

• 2-5. Embarked가 ‘C’인 승객 수 계산

1st. Embarked 열의 값이 'C'인 승객들을 필터링하기 위해 조건을 설정합니다

2nd. 조건에 맞는 승객 수를 계산하여 결과를 출력합니다

SAMSUNG SW ACADEMY FOR YOUTH

함께가요 미래로!
Enabling People

내일 방송에서 만나요!

삼성 청년 SW 아카데미