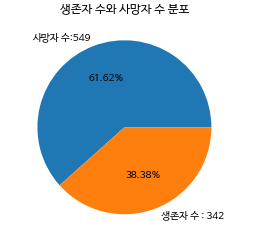
**빅데이터 수집/분석/시각화 과정 평가**

이름: 홍혁기

주어진 train.csv 파일을 이용하여 다음 문제를 해결하시오. 1번 18번은 10점, 나머지는 5점

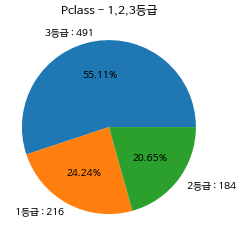
1. 전체 인원 수, 생존자 수, 사망자 수는 몇 명인가?

**전체 인원 수 : 891명, 생존자 수 : 549명, 사망자 : 342명**

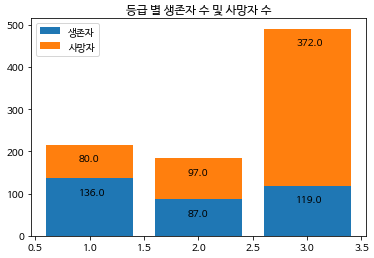
1. 생존자와 사망자를 Pie그래프로 나타내고 생존자와 사망자의 퍼센트를 소수점 2자리까지 나타내시오. (파이 이미지를 제출하시오.)
2. 결측 데이터가 있는 컬럼을 조사하시오.(컬럼 별로 결측 데이터 개수)

**Age : 177, Cabin : 204, Embarked : 2**

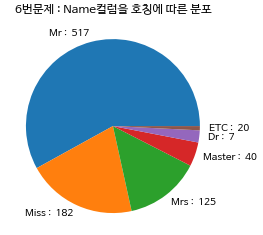
1. Pclass의 종류 별로 승객 분포를 파이 그래프로 나타내시오.



1. Pclass의 종류 별로 생존자 수와 사망자 수를 막대그래프로 나타내시오.(종류 별로 하나의 막대에 생존자 수와 사망자 수를 표시하시오.)



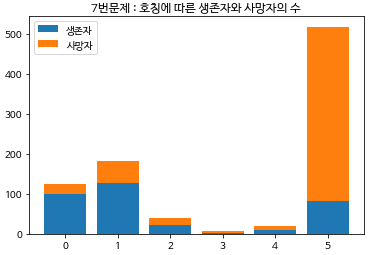
1. Name 컬럼에 호칭을 Mr, Mrs, Miss, Master, Dr, ETC(나머지)로 나누어 분포를 파이 그래프로 나타내시오.



1. 호칭에 따른 생존자 수와 사망자 수를 막대그래프로 나타내시오.(아래:생존자, 위:사망자)

**생존자 수 -> Mr : 81, Mrs : 99, Miss : 127, Master : 23, Dr : 3, ETC : 9**

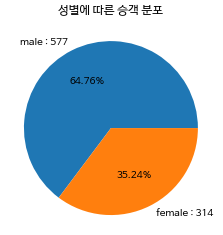
**사망자 수 -> Mr : 436, Mrs : 26, Miss : 55, Master : 17, Dr : 4, ETC : 11**



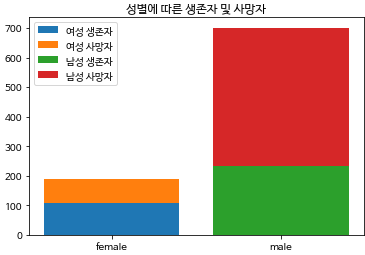
**X축 설명 -> 0 : Mrs 1 : Miss 2 : Master**

**3 : Dr 4 : ETC 5 : Mr**

1. 성별에 따른 승객 분포를 파이 그래프로 나타내시오.

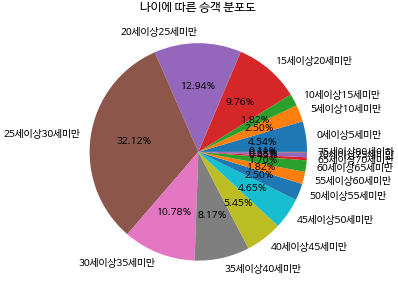


1. 성별에 따른 생존자 수와 사망자 수를 막대그래프로 나타내시오.(아래:생존자, 위:사망자)

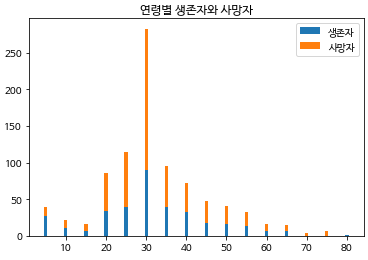


**여성 생존자 : 109, 여성 사망자 : 81, 남성 생존자 : 233, 남성 사망자 : 468**

1. 나이를 5살 범위로(0~5, 5~10,...) 나누어 승객 분포를 파이 그래프로 나타내시오.(결측값은 평균 나이로 처리하시오.)



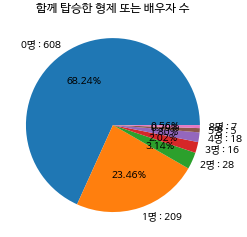
1. 나이(10번 문제에서 구분한 범위)에 따른 생존자 수와 사망자 수를 막대그래프로 나타내시오.(아래:생존자, 위:사망자)



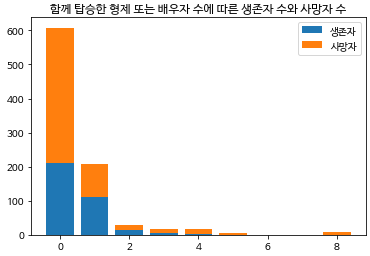
**생존자 수 : 27, 11, 7, 34, 39, 90, 40, 33, 18, 16, 14, 6, 6, 0, 0, 1**

**사망자 수 : 13, 11, 9, 52, 75, 193, 55, 39, 30, 25, 18, 10, 9 , 6, 0**

1. SibSp 값에 따라 승객 분포를 파이 그래프로 나타내시오.



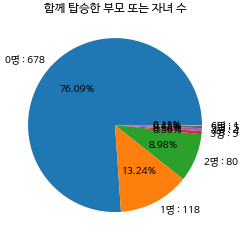
1. SibSp 값에 따른 생존자 수와 사망자 수를 막대그래프로 나타내시오.(아래:생존자, 위:사망자)



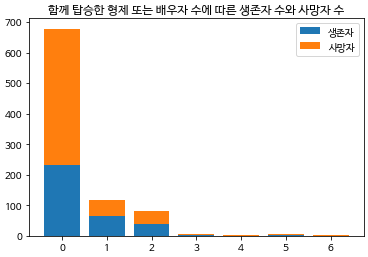
**생존자 수 : 0명 - 210, 1명 - 112, 2명 - 13, 3명 - 4, 4명 - 3, 5명 - 0, 8명 - 0**

**사망자 수 : 0명 - 398, 1명 - 97, 2명 - 15, 3명 - 12, 4명 - 15, 5명 - 5, 8명 - 7**

1. Parch 값에 따라 승객 분포를 파이 그래프로 나타내시오.



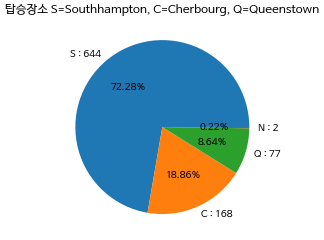
1. Parch 값에 따른 생존자 수와 사망자 수를 막대그래프로 나타내시오.(아래:생존자, 위:사망자)



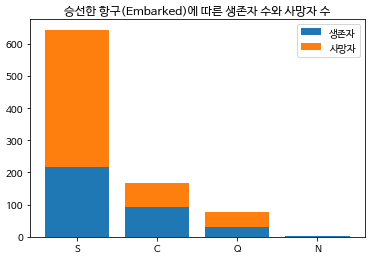
**생존자 수 : 0명 - 233, 1명 - 65, 2명 - 40, 3명 - 3, 4명 - 0, 5명 - 1, 6명 - 0**

**사망자 수 : 0명 - 445, 1명 - 53, 2명 - 40, 3명 - 2, 4명 - 4, 5명 - 4, 6명 - 1**

1. 승선한 항구(Embarked)에 따른 승객 분포를 파이 그래프로 나타내시오.(결측 값은 'N'으로 처리)



1. 승선한 항구(Embarked)에 따른 생존자 수와 사망자 수를 막대그래프로 나타내시오.(아래:생존자, 위:사망자)



**생존자 수 : S - 217명, C - 93명, Q - 30명, N - 2명**

**사망자 수 : S - 427명, C - 75명, Q - 47명, N - 0명**

1. Age 컬럼의 값을 표준 스케일로 변환하시오. (표준 스케일로 변환하면 평균은 0, 표준편차는 1로 바뀜, 결측치는 평균 값으로 처리)

from sklearn.preprocessing import StandardScaler

# 결측치 제거

age = data["Age"].fillna(data["Age"].mean())

# 2차원 배열로 변환

array = age.values.reshape(-1,1)

# 객체 생성

scaler = StandardScaler()

# 훈련데이터의 분포 저장

scaler.fit(array)

# 훈련 데이터 스케일링

scaler\_scaled = scaler.transform(array)

scaler\_scaled

