**빅데이터 수집/분석/시각화 과정 평가**

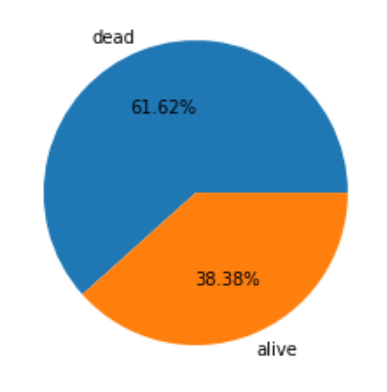
**이름:모범답안**

**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: 891 명 , 생존자 수:342명, 사망자 수:549명**

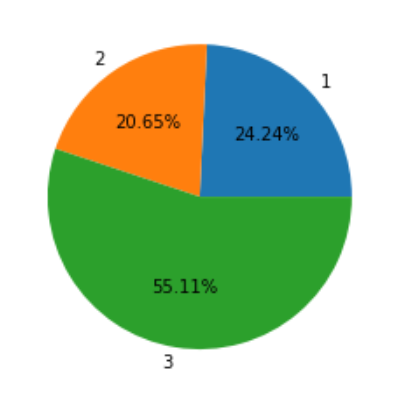
**2. 생존자와 사망자를 Pie그래프로 나타내고 생존자와 사망자의 퍼센트를 소수점 2자리까지 나타내시오. (파이 이미지를 제출하시오.)**

****

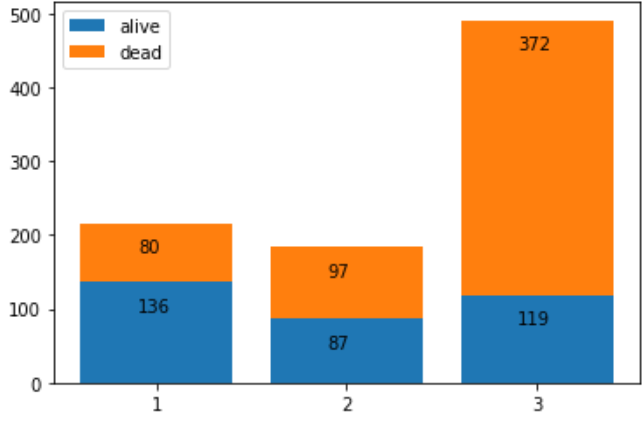
**3. 결측 데이터가 있는 컬럼을 조사하시오.(컬럼 별로 결측 데이터 개수)**

**"Age":177개, "Cabin":687개, "Embarked":2개**

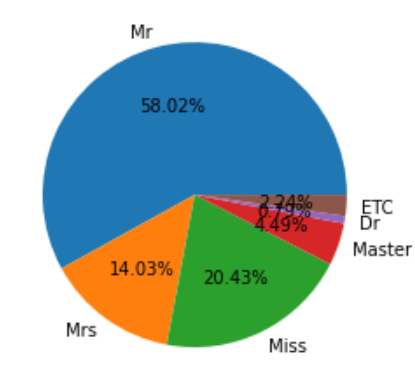
**4. Pclass의 종류 별로 승객 분포를 파이 그래프로 나타내시오.**

****

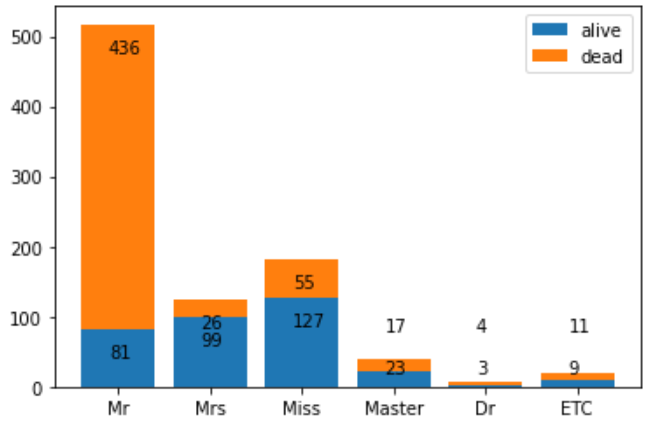
**5. Pclass의 종류 별로 생존자 수와 사망자 수를 막대그래프로 나타내시오.(종류 별로 하나의 막대에 생존자 수와 사망자 수를 표시하시오.)**

****

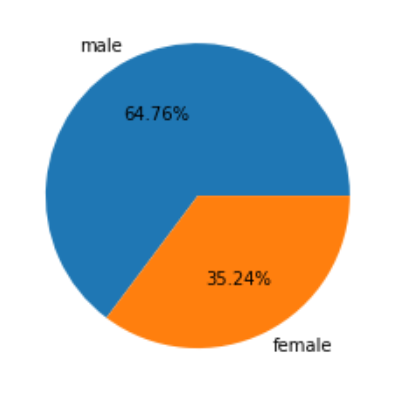
**6. Name 컬럼에 호칭을 Mr, Mrs, Miss, Master, Dr, ETC(나머지)로 나누어 분포를 파이 그래프로 나타내시오.**

****

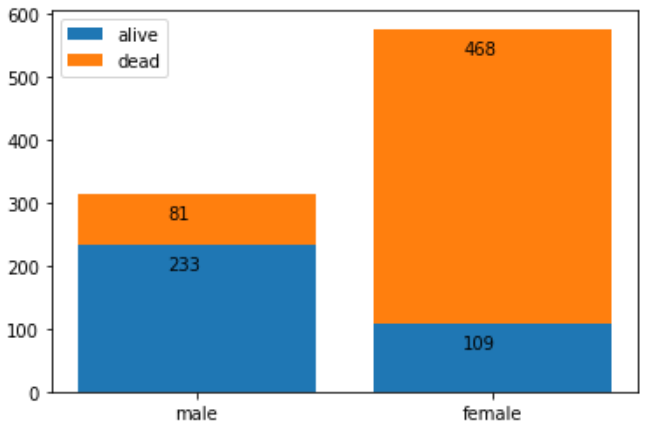
**7. 호칭에 따른 생존자 수와 사망자 수를 막대그래프로 나타내시오.(아래:생존자, 위:사망자)**

****

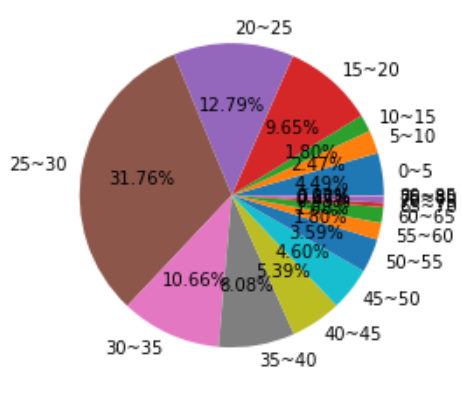
**8. 성별에 따른 승객 분포를 파이 그래프로 나타내시오.**

****

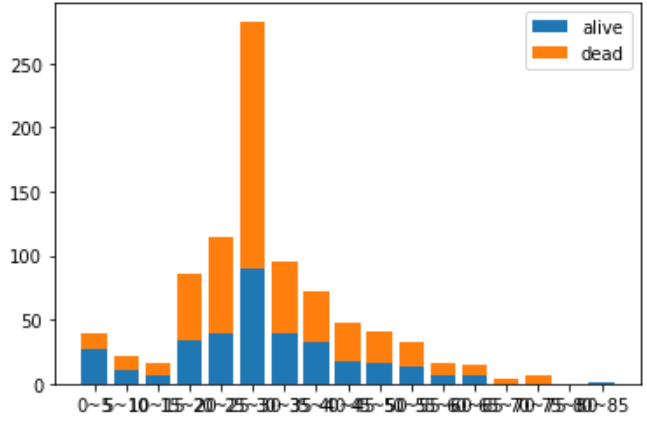
**9. 성별에 따른 생존자 수와 사망자 수를 막대그래프로 나타내시오.(아래:생존자, 위:사망자)**

****

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

****

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

****

00~05 27.0 , 13.0

05~10 11.0 , 11.0

10~15 7.0 , 9.0

15~20 34.0 , 52.0

20~25 39.0 , 75.0

25~30 90.0 , 193.0

30~35 40.0 , 55.0

35~40 33.0 , 39.0

40~45 18.0 , 30.0

45~50 16.0 , 25.0

50~55 14.0 , 18.0

55~60 6.0 , 10.0

60~65 6.0 , 9.0

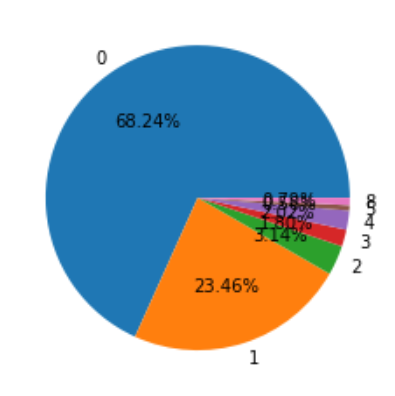
65~70 0.0 , 4.0

70~75 0.0 , 6.0

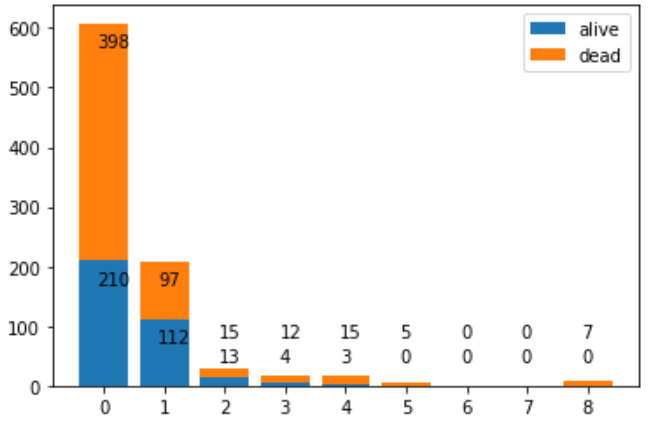
75~80 0.0 , 0.0

80~85 1.0 , 0.0

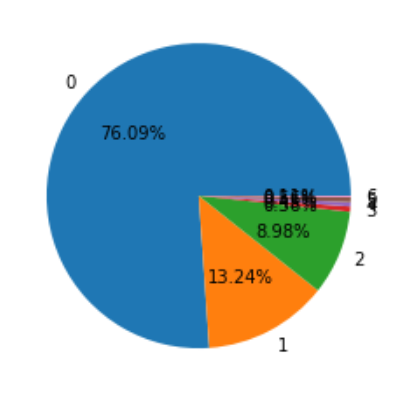
**12. SibSp 값에 따라 승객 분포를 파이 그래프로 나타내시오.**

****

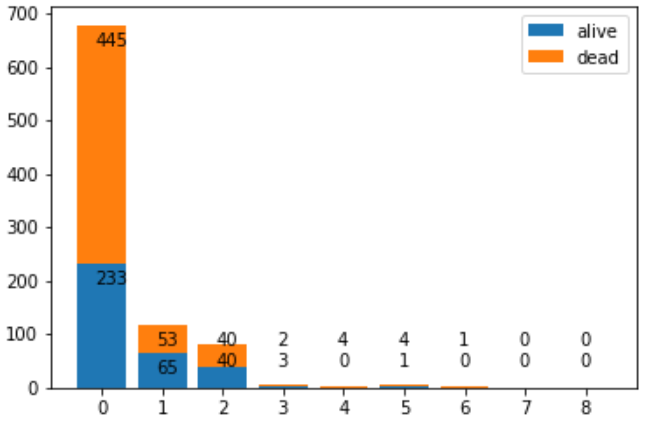
**13. SibSp 값에 따른 생존자 수와 사망자 수를 막대그래프로 나타내시오.(아래:생존자, 위:사망자)**

****

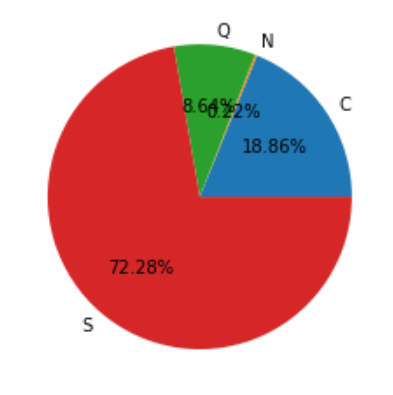
**14. Parch 값에 따라 승객 분포를 파이 그래프로 나타내시오.**

****

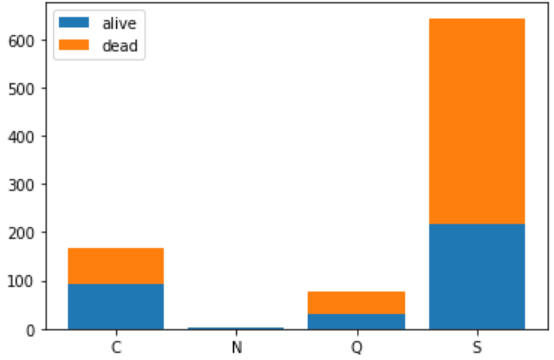
**15. Parch 값에 따른 생존자 수와 사망자 수를 막대그래프로 나타내시오.(아래:생존자, 위:사망자)**

****

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

****

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

****

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

a\_data = df['Age']

age\_mean = a\_data.mean()

age\_std = a\_data.std()

print("변환 전 평균:",age\_mean, " 표준 편차:",age\_std)

a\_data = (a\_data - age\_mean)/age\_std #표준 스케일 변환

age\_mean = a\_data.mean()

age\_std = a\_data.std()

print("변환 후 평균:",np.round(age\_mean,2), " 표준 편차:",age\_std)

변환 전 평균: 29.69911764705882 표준 편차: 14.526497332334042

변환 후 평균: 0.0 표준 편차: 1.0