

시각화 세션 과제

예상 독자

승선객 여행자 보험 설계자

시각화 목적

1. 보험 가입 시 입력하는 기본 정보인 연령, 성별과 생존 여부와의 관계 파악
2. 생존여부에 영향을 주는 다른 요소들과 연령, 성별과의 관계 파악
3. 타이타닉호의 승선객 사망자와 생존자의 분포 특성 파악

→ 위험도에 따라 보험 가격을 달리 설정할 수 있도록 함

시각화 Tool 선정 이유

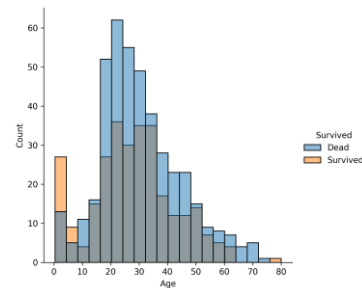
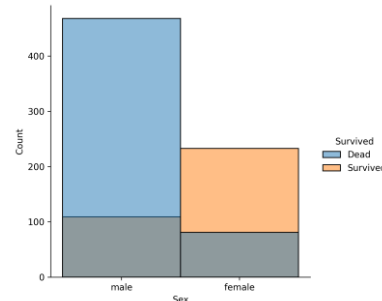
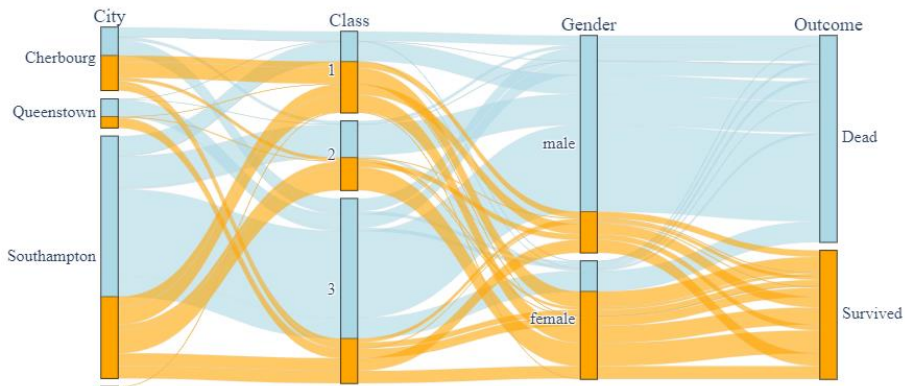
비교적 익숙한 tool이자

다양한 라이브러리를 활용하여 기본 그래프 ~복잡한 그래프까지 작성 가능

Plot 선정 이유

1. *displot*: 그룹별 차이를 쉽게 인식 할 수 있으며, 구분이 세세할 경우 그룹 내 빈도의 추이 또한 비교 가능함. → 주요 요소인 성별, 연령과 생존율의 관계를 한 눈에 알아 볼 수 있는 그래프 작성을 위해 선정함
2. *Parallel Categories Diagram*: plotly를 활용해 그릴 수 있는 그래프로, 다양한 요소의 상호 영향 관계를 보는 데 용이함 → 다른 영향 변수들과 주요 변수 간에 어떠한 관계가 있는지 대략적으로 보기 위해 선정함

시각화 자료 및 결과



생존자 비율

- ✓ 도시별 : Cherbourg > Queenstown > Southampton
- ✓ Class별 : 1 > 2 > 3
- ✓ 성별 : 남성 > 여성

승선지(출발지)별 탑승자 중 생존자 비율은 Cherbourg가 과반으로 가장 크며, Cherbourg 지역에서 승선한 사람들은 대부분 1등석 탑승권을 구매하였음. Class에 따라서는 1등석이 가장 생존율이 높았고, 성별에 따라서는 남성의 생존율이 여성보다 훨씬 낮았음. 연령별 차이는 뚜렷하지 않게 나타났고, 생존자 수의 빈도 또한 유사한 추이로 나타났음

→ Southampton에서 탑승한 남성 고객에게 할증을 두어야함