안녕하세요. 온라인 데이터 과학 커뮤니티인 Kaggle에 오신 것을 환영합니다. 대부분의 초보자들은 현장에서 길을 잃습니다. 왜냐하면 그들은 블랙박스 접근법에 빠지기 때문입니다. 그들이 이해하지 못하는 라이브러리와 알고리즘을 이용합니다. 이 튜토리얼은 데이터 과학자처럼 생각하는 방법과 생각하는 방법 및 코딩 방법을 알려줌으로써 동료보다 1-2년 앞서 나갈 수 있는 기반을 제공합니다. 첫 번째 대회를 제출할 수 있을 뿐만 아니라 앞으로 던져진 어떤 문제도 해결할 수 있습니다. 저는 명확한 설명, 깨끗한 코드, 그리고 풍부한 리소스 링크를 제공합니다. 참고하십시오. 이 커널은 아직 개선 중입니다. 업데이트 내용은 아래의 변경 로그를 확인하십시오. 또한 상향 투표와 포크, 코멘트도 꼭 부탁드리며 앞으로도 발전해 나가겠습니다. 고마워요, 그리고 "통계적으로 의미 있는" 행운이 있길 바래요!

**1단계: 문제를 정의합니다.**

이 프로젝트를 위해, 문제점은 금도금 위에 우리에게 주어지고, 타이타닉호의 승객들의 생존 결과를 예측할 수 있는 알고리즘을 개발합니다.

프로젝트 요약: RMS 타이타닉호의 침몰은 역사상 가장 악명 높은 난파선 중 하나입니다. 1912년 4월 15일, 첫 항해 중, 타이타닉호는 빙산과 충돌한 후 가라앉아 승객과 승무원 2224명 중 1502명이 사망했습니다. 이 선풍적인 비극은 국제 사회에 충격을 주었고 배를 위한 더 나은 안전 규제로 이어졌습니다.

난파선이 이처럼 인명피해를 초래한 이유 중 하나는 승객과 선원들을 위한 구명보트가 충분하지 않았기 때문입니다. 비록 침몰에서 살아남는 행운의 요소가 일부 있었지만, 여성, 어린이, 그리고 상류층과 같은 몇몇 사람들은 다른 사람들보다 살아남을 가능성이 더 높았습니다.

이 도전에서는 어떤 종류의 사람들이 살아남을 가능성이 있는지에 대한 분석을 완료해 주시길 부탁드립니다. 특히 어떤 승객이 비극에서 살아남았는지 예측할 수 있도록 머신러닝의 도구를 적용해 주시길 부탁드립니다.