

## Youtube 广告投放决策

### 项目简介:

你是一名数据科学家 (DS, senior BA, senior DA) , 受雇于一家上市公司。

### Main Project:

由于公司今年 (2017 年) 销售额增长放缓, 高层决定通过在视频平台 Youtube 上进行广告投放来增加销售额。现在, 项目经理需要你出具一份数据分析报告并在高层会议上为同事进行解读, 帮助公司做出以下决策:

- 应该对哪些视频进行投流
- 应该在什么时间点上进行投流
- 为应对即将到来的销售额增长, 公司应该做什么规模的准备

### After Project:

投放完成后, 项目经理希望你对公司销售额进行评估, 从而判定本次广告投放是否取得了预期的效果。

---

Tips: 为了完成本任务, 你可能需要做如下工作:

1. 收集数据并进行基本的数据处理
  2. 根据数据, 决定你的数据分析策略
  3. 依据分析策略, 建模并得出分析结论
  4. 根据分析结论, 制作讲演 PPT, 将重要的数据分析成果绘制成图, 并进行展示
  5. 进行数据分析结果 presentation
  6. 使用 ABTest 进行广告投放的效果验证
- 

### MLE Plus(for MLE)

我们知道, 视频网站上, 点赞和评论越多的视频, 其越容易被更多的人看到。基于这样的商业广告投放环境, 项目经理认为, 要想提高本公司广告投放的效果, 可以通过在投放了公司广告的视频下方进行评论的方式来实现。但是, 人工进行评论的时间成本过高, 因此, 项目经理希望你运用机器学习的方法, 自动生成对广告和视频内容的评论, 并添加到视频下方。

---

Tips: 为实现 MLE 目标, 你可能需要:

1. 收集评论的文本数据并进行数据处理

2. 选择合适的机器学习模型
  3. 进行模型训练
  4. 实现模型部署
  5. 调用该模型接口，并实现自动追加评论
- 

本项目为解决上述这样一个真实的互联网商业决策问题而设计，采用真实的 Youtube 视频网站数据。

在数据分析工作流程方面中，我们按照数据分析的流程，对 Youtube 网站相关数据进行数据数据清洗，数据分析，数据建模，数据预测，最终实现商业决策。并使用 hold out 数据进行决策结果的验证。

在数据表达方面，我们会使用得到的结论进行基于 tableau 的可视化表示，帮助学生进行 PPT 准备并模拟进行 presentation。

在技术栈方面，我们将用到经典的回归，分类和聚类模型，实现对视频网站数据的有效处理，对于 MLE 学生，我们可以采用 RNN 等模型实现文本生成。

#### 项目关键词:

数据分析、数据科学、Predict, Clustering, 预测模型  
机器学习，自然语言处理，NLP,  
难度：易、中、进阶均可，可根据学员吸收情况进行调整

#### 项目需要具备的背景知识:

Python, 统计、数学、建模有最基本的了解

#### 项目需要用到的工具或软件:

SQL  
Python (Numpy, Pandas, Scikit-Learn, Matplotlib, etc.)  
Tableau  
Microsoft Office  
AWS  
GitHub

#### 项目最终的目标和成果:

1. 真实的商业智能 (BI) 决策经验
2. 数据分析的标准流程
3. 可展示的项目成果

Project Outline:

*以下项目大纲仅供参考，可实时按需进行调整。*

## **Module 1: Kick off**

1. Project Intro
2. Student presentation
3. Reach a Consensus on the Project
4. Achieve Overall Project Planning for student

## **Module 2: Data Preprocessing**

1. Environment Installing (python,sql)
2. Data Viewing
3. Data Cleaning
4. Data Transformation
5. Featuring

## **Module 3: Data Analysis and Statistics**

1. Determine data analysis strategy
2. Conduct statistical level data analysis
3. Draw fundamental conclusions
4. Determine direction for further analysis

## **Module 4: Predictive Task Modeling**

1. Data Modeling Methodology
2. Linear Regression, Decision Tree and Gradient Enhancement method
3. Modeling and predict Youtube Video views, Likes and Comments

4. Identification of Fake Video Data
5. Model optimization
6. Model evaluation and comparison

## **Module 5: Cluster Task Modeling**

1. Unsupervised Learning
2. K-Means Algorithm
3. Clustering and classify source data
4. Analysis of data clusters

## **Module 6: Data visualization**

1. Visualization by Excel
2. Visualization by Tableau
3. Visualization by Python

## **Module 7: Knowledge Point Go Through**

1. DS knowledge point go through
2. MLE knowledge point go through

## **Module 8: Classification Task Modeling(MLE)**

1. SVM,Random Forest,Neural Network method
2. Modeling and classify Youtube video
3. Hold out test data and model evaluating
4. Model optimizing

## **Module 9: Model Deploying(MLE)**

1. AWS Intro

2. AWS utilizing
3. Model deploying on AWS

## **Module 10: Project Outcome Production**

1. Analysis result review
2. Business story telling
3. PPT creation guidance
4. Business presentation guidance

## **Module 11: Resume**

1. Project experience
2. Add This Project into Resume
3. Resume optimization
4. Github utilizing
5. Upload this project to github

## **Module 12: Presentation & Mock Interview**

1. Presentation of the project
2. Mock Interview