

프로그램 북

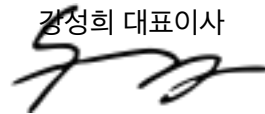
DS School 수업 소개

실전 데이터분석반

수강료 : 594,000원

DS School (더넥스트스쿨)

강성희 대표이사



강연 진행

강연은 총 4주간 진행되며, 주 1회씩 총 4회를 진행합니다. 매 회차마다 실습이 포함되어 있으며, 이를 위해 회당 강연 시간은 다섯 시간으로 책정하였습니다. 그러므로 총 수업 시간은 4 회 20 시간입니다. 강연의 세부 내용은 다음과 같습니다.

실전 데이터분석반 1주차

데이터 분석의 기본을 배우고, 헬스케어 스타트업의 회원 정보와 행동 데이터를 정리하고 분석하기

첫 수업에서는 데이터 분석의 기본을 배웁니다. 데이터 분석(Data Analytics)의 의미와 역할, 데이터 분석을 통해 얻을 수 있는 것을 배운 뒤, 파이썬의 데이터 분석 패키지 판다스(Pandas)를 학습합니다.

판다스(Pandas)를 사용하면 대용량의 데이터를 효율적으로 분석할 수 있고, 복잡한 분석도 파이썬을 활용해 간단하게 처리할 수 있습니다. 수업에서는 판다스의 가장 기본적인 데이터 형식인 데이터프레임(DataFrame)의 사용법부터, 데이터에서 행렬을 가져오고, 색인하고, 정렬하고, 잘못된 데이터를 정리하는 법 등을 배웁니다. 또한 피벗 테이블(pivot_table)과 같은 간단한 분석 노하우도 배울 수 있습니다.

판다스(Pandas)를 모두 배웠으면, 헬스케어 스타트업 نوم(Noom)의 데이터를 활용해 분석을 시작합니다. 먼저 잘못된 회원 정보(몸무게가 너무 높거나, 키가 너무 작거나)를 정리하는 데이터 클리닝(Data Cleaning)을 진행한 뒤, 다음의 질문에 대해 데이터에 기반한 답을 제시합니다.

- نوم(Noom)을 사용하는 무료 사용자들은 주로 어느 시간대에 유료 사용자로 전환하는가? (ex: 새벽, 낮, 주중, 주말) - 이 사실을 알 수 있다면, 해당 시간대에 유료 서비스 구매를 촉진하는 이메일을 보냄으로써 회사의 매출을 올릴 수 있습니다.
- نوم(Noom)을 사용하는 무료 사용자 중, 유료 사용자로 전환할 확률이 가장 높은 연령/성별은 어디인가? - 이 사실을 알 수 있다면, 해당 연령/성별에 집행하는 마케팅 예산을 늘릴 수 있습니다. (가령 30대 여성이 유료 결제 확률이 높다면, 페이스북에서 30대 여성에게 광고를 더 많이 보여줘야겠죠) 반면 정 반대의 경우에는 마케팅 예산을 줄임으로써 더 효율적인 마케팅을 할 수 있습니다. (가령 50~60대 남성은 유료 결제 확률이 낮다면, 페이스북 마케팅에서 이분들에게는 광고를 노출하지 않는 것도 좋은 방법입니다)
- 유료 사용자를 코칭하는 코치 중, 어느 코치가 가장 만족도가 높은가? - 만족도가 높은 기준은 사용자가 계속 유료 서비스를 이용하는가로 판단할 수 있습니다. 즉, 고객 이탈률(churn rate)이 가장 낮은 코치가 가장 뛰어난 코치라고 판단할 수 있습니다. 반면, 고객 이탈률(churn rate)이 높은 코치일수록 사용자가 가장 불만족스러워한다고 판단할 수 있습니다.

실전 데이터분석반 2주차

직무교육 스타트업의 마케팅 정보와 구매 정보를 통해 효율적인 마케팅 채널을 찾아내기

두 번째 수업에서는 여러 개의 데이터를 하나로 합쳐서 사용하는 법과 데이터베이스에서 필요한 정보를 가져올 방법을 배웁니다. 대부분의 경우, 데이터는 하나의 파일(excel, csv, etc)로 저장되어 있지 않습니다. 효율적으로 데이터를 저장하기 위해 여러 개의 작은 파일로 나누어져 있으며, 데이터를 분석할 때는 이렇게 여러 개로 나뉜 데이터를 하나로 합치는 방법을 배워야 합니다. 또한 데이터는 CSV와 같은 파일 형식이 아닌 MySQL, MSSQL, Oracle과 같은 전문적인 프로그램(일명 데이터베이스)에 저장된 경우도 있습니다. 이 경우 SQL(Structured Query Language)이라는, 데이터베이스를 위한 별도의 프로그래밍 언어를 사용해야 합니다.

이번 수업에서는 판다스에서 여러 개의 데이터를 하나로 합치는 merge, join, concat 등의 기능, 그리고 데이터 분석가에게 필요한 SQL 문법을 압축적으로 배울 것입니다.

여러 개의 데이터를 합치는 법을 배웠으면, 데이터사이언스 교육 스타트업 DS School의 마케팅 데이터와 결제 데이터를 분석합니다. 분석 목표는 어느 마케팅 채널이 가장 효율적이며, 반대로 어느 마케팅 채널이 가장 효율적이지 않은가? 입니다. 효율적인 마케팅 채널이 있다면 해당 채널에 마케팅 예산을 집중하고, 반면에 효율적이지 않은 채널이 있다면 해당 마케팅 채널에 집행한 예산을 줄일 수 있습니다. 마케팅 효율에 대한 구체적인 질문은 다음과 같습니다.

- DS School의 수강생은 어떠한 목적을 달성하기 위해 수업을 신청하였는가? - DS School 수강생의 구매 전 설문 조사를 바탕으로 이 내용을 분석할 수 있습니다. 이 분석 결과를 바탕으로 수업의 내용을 변경하거나, 난이도를 조정하거나, 수강생의 니즈에 맞는 신규 커리큘럼을 만드는 것도 가능합니다.
- DS School이 보유한 마케팅 채널 중 가장 효율이 높은 채널은 어떤 것인가? - DS School은 주로 페이스북 마케팅을 활용하는데, 이 채널당 고객 획득 비용(Customer Acquisition Cost, 이하 CAC)과 고객 생애 가치(Customer Lifetime Value, 이하 LTV)를 계산한 뒤 LTV/CAC 비율이 가장 효율적인 마케팅 채널 순으로 정렬합니다. 가장 효율이 높은 채널을 파악할 수 있다면, 이 채널에 예산을 집중하거나 이 채널이 효과가 좋은 이유를 분석해서 새로운 마케팅 채널과 콘텐츠를 생성할 수 있습니다.
- DS School의 매출 변화는 어떻게 되는가? - DS School의 강연이 가장 잘 팔리는 시기(시간, 내지는 주중/주말 등)를 파악한 뒤 그 시기에 구매를 유도하는 이메일을 발송할 수도 있고, 현재의 매출 변화량을 바탕으로 누적 매출이나 예상 매출 등을 리포트 방식으로 시각화시킬 수도 있습니다.

실전 데이터분석반 3주차

여성 의류 쇼핑몰 플랫폼의 고객 행동 패턴을 분석하여 서비스의 매출을 증대시키기

세 번째 수업에서는 앞서 배운 판다스와 SQL 노하우와 더불어, 데이터를 시각화한 뒤 이 결과를 바탕으로 결론을 도출하는 방법을 배웁니다.

데이터 시각화는 주어진 데이터를 직관적으로 표현할 수 있다는 장점이 있으며, 그렇기 때문에 분석 결과를 다른 사람에게 이해시키고 설득할 수 있도록 도와줍니다. 하지만 효과적이고 직관적인 데이터 시각화를 위해서는, 이를 위한 적절한 시각화 그래프나 도표를 사용해야 합니다.

이번 수업에서는 파이썬의 데이터 시각화 패키지인 seaborn이나 matplotlib등을 활용하여 데이터를 분석하고 시각화 할 수 있는 방법을 배웁니다. 이 방법을 활용하면 파이썬과 판다스, 주피터 노트북으로 직관적이고 편리한 시각화 대시보드를 구현할 수 있습니다.

데이터 시각화를 배웠다면, 다음으로는 국내 최대의 여성 의류 쇼핑몰 플랫폼 지그재그(ZIGZAG)의 고객 행동 패턴을 분석합니다. 이번 목표는 사용자의 액티비티를 바탕으로 고객의 니즈나 주요 여성 의류 쇼핑몰들의 특징을 분석하는 것입니다. 또한 이번 수업에서는 데이터 분석가를 위한 분석뿐만 아니라, 기획자, 마케터, 내지는 팀 내의 의사결정권자에게 데이터를 직관적으로 이해시킬 수 있는 시각화 방법을 배울 것입니다. 구체적인 분석은 다음과 같습니다.

- 고객들은 지그재그에서 어떠한 과정을 거쳐서 상품을 구매하는가 - 이러한 행동 패턴을 분석하면 고객의 구매 루트를 최적화하거나, 구매하지 않고 이탈할 사용자를 붙잡아서 매출을 증대할 수 있습니다.
- 고객들이 선호하는 여성 의류 쇼핑몰들은 어디인가? - 이 사실을 파악할 수 있으면, 고객이 서비스를 이용하는 과정 중에 (내지는 이용 후에 이메일을 통해) 쇼핑몰이나 상품을 추천한 뒤 고객에게 상품을 업셀(up sell, 추가로 구매하도록 유도)할 수 있습니다.

실전 데이터분석반 4주차

프리랜서 오픈마켓 서비스의 사용자 로그 데이터를 분석하여 사용자의 행동 패턴을 파악하기

마지막 수업에서는 기존에 배운 내용 외에도, 수강생분들이 앞으로 데이터 분석가로서 활동할 때 알아둬야 하는 부분들을 다룹니다. 판다스에서 데이터를 효율적으로 읽고 저장하는 방법, 텍스트 파일(특히나 한글 데이터)를 읽다가 문제가 발생할 경우 이를 해결하는 방법, 대용량의 데이터를 효율적으로 분석하고 저장하는 방법 등에 대해 집중적으로 배웁니다.

마지막 수업에서는 그동안 배운 내용을 총동원하여, 한 명의 데이터 분석가로서 수강생이 어느 정도의 실력을 갖추고 있는지 검증하는 시간을 가질 것입니다. 이번 수업에서는 프리랜서 오픈마켓 서비스 크몽(Kmong)의 사용자 데이터를 분석할 것입니다. 주요 분석 과제는 다음과 같습니다.

- 크몽 사용자의 웹 행동 데이터, 그리고 앱 행동 데이터를 정리하기 - 크몽의 고객들은 웹, 아이폰, 안드로이드 등 다양한 디바이스와 OS로 서비스를 이용하고 있습니다. 이 디바이스에서 쌓인 내용을 분석할 수 있도록 깔끔하게 정리한 뒤, 여러 개로 나누어진 데이터를 판다스의 merge나 concat등을 활용해 하나로 묶어주는 작업을 할 것입니다.
- 크몽 사용자들의 유입 경로, 검색 키워드를 분석하기 - 크몽의 고객들은 네이버나 페이스북 등 다양한 경로를 통해 서비스를 방문합니다. 이 사용자들이 어떤 채널(네이버, 페이스북, etc)을, 어떤 키워드나 콘텐츠를 통해 방문할는지 분석을 할 수 있다면 마케터들이 신규 광고 콘텐츠를 만들거나 검색 키워드를 늘리는데 도움을 줄 수 있을 것입니다.
- 크몽 사용자들의 구매 루트 분석하기 - 크몽의 고객들이 서비스에 유입되었다면, 무언과의 과정을 통해 콘텐츠를 살펴보고, 검토한 뒤 구매를 결정했을 것입니다. 이 과정에서 고객이 구매를 결정할 수 있었던 요소를 분석할 수 있다면, 서비스 차원에서 이를 더 강조하거나 부족한 부분을 충원한다면 고객들이 구매를 결정하는 데 도움을 줄 수 있고, 이는 서비스의 매출 증대로 이어질 수 있을 것입니다.

실전 데이터 분석반 수업 이후

주어진 과제를 푼 뒤, 이력서와 함께 제출하기

4주간의 수업이 모두 끝나면, 여태까지 분석한 콘텐츠 외의 추가 과제가 수강생들에게 제공될 것입니다.

추가 과제는 3주차 수업인 지그재그 데이터와 4주차 수업인 크몽의 데이터를 활용한 각각 10개의 추가 과제가 주어집니다. 만일 데이터분석가로 취업하거나 이직하길 원하는 분들이 있다면, 해당 과제를 푼 뒤 이력서와 함께 제출하면 관련 자료가 각각의 스타트업으로 전달됩니다.

전달 후 스타트업 담당자분의 검토를 거쳐, 데이터분석가로서 활동할 수 있는 충분한 소질이 있다고 판단될 경우 본격적으로 면접 진행 절차를 밟게 될 것입니다.