Case Study Preview

Data Wrangling using R: Application to Netflix Data

Unnati Narang, Assistant Professor, Business Administration, Gies College of Business

Published Sep 1, 2022

> Teaches students about the basics of ETL (extracting, transforming, and loading data) or data wrangling (e.g., loading a dataframe, subsetting data, etc.) in R.

About this case study

This case teaches students the basics of ETL (extracting, transforming, and loading data), or data wrangling, in R: loading a dataframe, subsetting data, data discovery, data structuring and data cleaning. The case questions are solved using a dataset on TV shows and movies from Netflix, loaded and run in R.

It is suited to novices in advanced undergraduate or graduate classes in any business discipline. In a suggested first lecture, the instructor walks through the slides, helps students install R and RStudio, demonstrates R with the demo files, and introduces the Netflix case; a second lecture discusses the solution.

At a glance

Type
Case Study

Audience
Faculty, Student

Discipline
Business

Skills
Data Analysis
Extract Transform Load

Software
R

Get the full case study, free with an account: **emergingtech.giesbusiness.illinois.edu/?p=481**

A look inside

The opening of the case, as students read it.

Data Wrangling using R and its Application to Netflix Data

Introduction

The volume of data generated and consumed worldwide is growing rapidly. Consider how much data you generate daily as a consumer. Every time you interact with your computer or smartphone or make a transaction, new data are generated. Every social media engagement generates data. Every instance you drive or walk outside with your phone generates data about your location in the form of high-frequency GPS, or global positioning system and tracking.

The rapid rise in the volumes and types of data due to the digital revolution has become deeply embedded in many facets of our lives. The total volume of data created, recorded, copied, and consumed around the world in 2018 was 33 zettabytes (ZB), or 33 trillion gigabytes. This increased to 59ZB in 2020, and by 2025, it is expected to reach a mind-boggling 175ZB. Often these increasing volumes of data can contain valuable information and insights to help researchers and managers make better decisions.

How can you learn to make sense of data? How can you identify patterns in the data that could reveal important insights? How can you use R programming to make the process of data wrangling more efficient? Because data analytics is frequently used to inform business and policy decisions, data wrangling must be done carefully and thoroughly. The process of cleansing and integrating chaotic and complicated data sets for easy access and analysis is known as Data

Preview ends here. The full case goes on to R, the Netflix data and the case questions.

What is included

To share with students

- Case document (Word)
- Netflix data (CSV) and data description sheet

For instructors only

- Solution in R Markdown, with an HTML version
- Lecture slides and an R demo with data
- Read-me (Word)

Get the full case study

Free with an account on the Center website. Scan the code or visit

emergingtech.giesbusiness.illinois.edu/?p=481

For instructors

Choose Instructor when you sign up. Instructor materials are shared once your instructor status is verified.