

Data Science Capstone Project: Movie Recommendations

Catherine Edis

9 January 2022

Introduction

Purpose

This project was undertaken to fulfill the requirements of the “Data Science: Capstone” course, which is part of the Professional Certificate in Data Science program offered by Harvard University. The purpose of the project was to develop a movie recommendation system using the tools explained throughout the courses in the Professional Certificate program.

The system was developed using the “MovieLens” movie ratings dataset published at: <https://grouplens.org/datasets/movielens>

MovieLens

MovieLens is a non-commercial service operated by GroupLens Research at the University of Minnesota. It invites its users to rate movies, and uses “collaborative filtering” technology to provide personalized movie recommendations.

The GroupLens Research Group periodically publishes the MovieLens dataset for University students to research personalization, filtering and prediction technologies. The full dataset was last published in September 2018 and contained approximately 27 million ratings. A smaller, stable dataset - the “MovieLens 10M dataset” - was used for this project. It was released in January 2009, and contains approximately 10 million ratings. The MovieLens 10M dataset is available at: <https://grouplens.org/datasets/movielens/10m>

[Describe the data.]

Goal of the Project

[Measure/aim.]

Key Steps Performed

[Describe them.]

Analysis

[Explain the process and techniques used:]

Results

[Present the modelling results and discuss the modelling performance.]

Conclusion

[A brief summary of the report, its limitations and future work.]

Appendices

R Markdown - Remove these sections before submitting

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

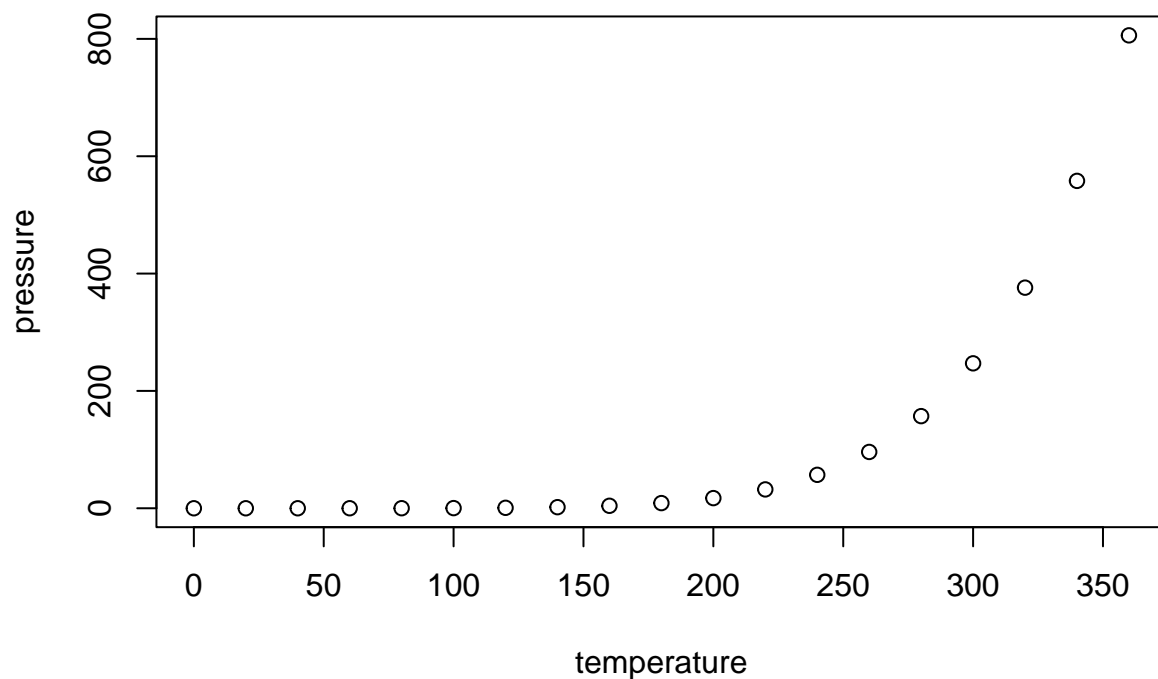
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean    : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.    :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.