Final Project of Introduction to Data Science Fall 2020
White Turnips Data Analysis

Introduction

White turnips are a unique type of flora present in every Animal Crossing series game to date. They can be bought from Joan or Daisy Mae (Joan's granddaughter) every Sunday morning before 12 p.m. Unlike other items, the price of turnips fluctuates over time. Joan will have a random buying price each Sunday and over the week, the local store will buy them at different prices. These prices will differ each morning and afternoon. This trade is nicknamed the Stalk Market. Turnips must be sold within a week, otherwise they will rot by the following Sunday, and lose their monetary value.

Source: `https://animalcrossing.fandom.com/wiki/White_turnip`

Data Type

All buying prices on Sunday and selling prices of morning and afternoon from Monday through Saturday were collected from 4/26/2020 to 7/18/2020. Data set is available under the "Final Project" folder in Canvas.

Goals of the Final Project

10 questions will be listed under the "question" section and you have to work on all of them by analyzing the data set.

Requirement

(1) You have to use R to analyze the data set.
(2) Borrow any information from the existing website for turnip prediction is not allowed, please analyze the data set and generate all the results by yourself.
(3) Please generate your final report using R markdown you learned from this class. And submit both your final report and Rmd format files through Canvas.
(4) You are expected to work independently. **Offering** and **accepting** solutions from others is an act of **plagiarism**, which is a serious offense and **all involved parties will be penalized according to the Academic Honesty Policy**.
(5) Although this is the game data, I do not encourage you play or buy this game. First, Nintendo did not pay me to advertise this game, second, playing this game will not help you work or getting better quality of the final project at all.

Rubric

Final project contains 30% of your final grade, and there are 10 questions with relative hypothesis of this project (3 points for each question). There is no unique soultion for each question, please study the data structure, and think about how to generate the result and interpret your result base on your own knowledge and everything you learned from this course. Using the fancy or complicated method will not getting better grade for this project, what I need is reasonable method to generate the result and you are able to interpret your finding for each question.

I will strictly follow the following rubric for grading of your final report, for each question (3 pts):

(1) (0.5 pt) Brief describe the method you are going to use to answer the question, show the formula and cite any paper if necessary.
(2) (0.5 pt) Explain why this method you use for this question is appropriate for answering this question.
(3) (0.5 pt) Show your R code (like the lecture note showing the code then output).
(4) (0.5 pt) Show your result generated by your R code (like the lecture note showing the code then output).
(5) (0.5 pt) Explain your result or finding.
(6) (0.5 pt) Summariz your result or make conclusion.

Missing any one of the above requirements will get 0 for that part.

Bonus Credit

If you can generate the data set into the data structure such as mpg, diamonds or flights as we used for this course (Not covered in the lecture but you can learn how to do this from textbook R for Data Science) and generate all of your results using ggplot function. You will get 20% bonus of your grade for your final project.

Please provide R code before the first question.

Questions

Please review carefully again of the rubric for the final project before you work on the following questions.

(1) Please find a good way to visulaize this data set in order to present better description of this data set using figures.

(2) Base on all buying price from the data set, can you estimate mean and the range of the white turnip buying price?

(3) Base on all selling price from the data set, can you estimate mean and the range of the white turnip selling price?

(4) Will the white turnip selling price in the afternoon significantly higher than the price in the morning?

(5) The white turnip selling prices on 5/1 is missing, this is because the Stalk market will be close for upgrading one day if the player play this game and open the market for 30 days. So selling price for upgrading is unavailable for one day. Can you find a way to impute reasonable number for the missing selling prices for morning and afternoon?

(6) From the game rule of introduction, we know that white turnips must be sold within a week, otherwise they will rot by the following Sunday. So that means the buying price will only affected by the selling price for the following 6 days. Base on the data set we have, what is the probability the player will earn money for buying the white turnip in this game? Please describe your way in detail to answer this

question and showing your result using table or figure to support your finding.

(7) If we call the buying price on Sunday and all the following selling prices from Monday through Saturday a period, Is there any specific pattern for selling price for all periods? If there exist patterns, then how many? (If you google the white turnip for this game, it is not hard to find the answer, and I can tell you there are four patterns for selling price, but you have to use your own way to find patterns).

(8) Now we know there are four different patterns for selling prices every week, will the specific pattern for this week affect the probability of certain patterns for the following week? For example, if this week happens with pattern number 1, what is the probability of the other three patterns next week. Please answer this question by using the results (patterns) from previous question and use as underlying patterns.

(9) Will certain days for selling price from Monday to Saturday are tend to be higher or lower than all the other days? (For example, the selling price for all Monday from the data set are significantly lower than the selling price in other days, or the selling price for all Thursday are significantly higher than the selling price in other days). Please investigate the hypothesis and show your finding.

(10) If the goal of this project is to predict the selling price in order to earn a lot of money or not losing any money in this game, can you predict the selling prices from Thursday to Saturday if we already have the buying price on Sunday and all selling prices from Monday to Wednesday? Try to predict the price using two following independent scenarios:

Scenario 1: Buying price is 93 on Sunday, and the selling prices are 140 and 127 on Mondays, 183 and 212 on Tuesday and 158 and 83 on Wednesday?

Scenario 2: Buying price is 107 on Sunday, and the selling price are 104 and 138 on Mondays, 65 and 58 on Tuesday and 109 and 101 on Wednesday?

You can rely on some machine learning algorithm or modeling procedure for prediction of the selling prices.