E-14a

HW 1: Instructions

Rossmann Drugstore

Learning Objectives

- Learn how to use Python notebooks
- Use basic data modeling techniques
- Combine Python notebook with Flask
- Deploy your app using Heroku
- Create Rosmann app for data analysis

ROSSMANN PROJECT OVERVIEW

In September of 2015, a competition was launched to see which Data Science team could generate the most accurate predictions for the next six weeks of sales for each of the 1,115 Rossmann drug stores located throughout Germany. With a bounty of $30,000 in prize money for the winners, this challenge attracted some of the best researchers on the Kaggle Data Science website. That competition has since completed, but now provides a rich data set and research studies of some of the best modeling techniques currently in use.

Homework 1 description

This homework includes many sequential steps. Please, read the instructions carefully. Go to *Canvas>Modules>Homework 1* and download ".zip" file called *hw_1*. Unzip your folder and follow the next steps.

Part 1: Python notebook (5 pts)

In the first part of your Homework you will work with Python notebook. Open your terminal, navigate to your *hw_1* folder and run:

```
ipython notebook rossmann.ipynb
```

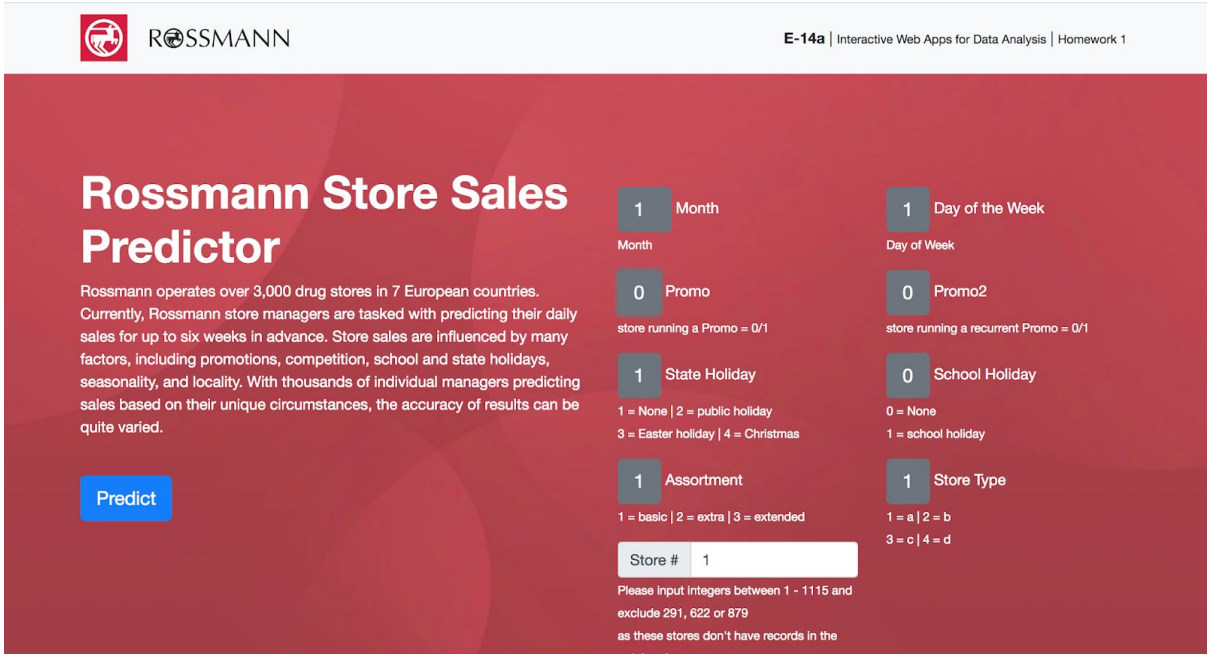
Your notebook will open in the browser. Read and accomplish all tasks listed in the notebook. You will be challenged to add some codes and answer some questions. Remember that skipping some parts may affect the rest of the notebook.

The objective of Part 1 is to help you learn the basic data science modeling and preparing the best model for your flask app.

Part 2: Flask App (5 pts)

We will use the best model from Part 1 to predict sales based on user's input. In this app user's will be able to choose the values from drop-down menus and based on its input users can obtain different prediction values.

This is how the final app should look like:



Preview of the Homework 1 interface

What we need to create in this part is:

- HTML: create drop-downs that will pick up the user's choices
- Python Flask: Use inputs as features and generate predictions (Complete implementation in the make_prediction() function in `routes.py`)

YOUR TURN:

Part 2.1: Implement bootstrap dropdowns in `index.html`

- Month,
- Day of the Week
- Promo,
- Promo2
- State Holiday,
- School Holiday
- Assortment,
- Store Type

Here is the example of a dropdown (style: refer to the screenshot):

```
<select name="example_name" class="example_class">

<option>1</option>
<option>2</option>
      ...
</select>
```

Part 2.2: Load ML model into Flask app

Put your best pickled machine learning (ML) model into your Flask app.

Part 2.3: Complete `make_prediction()` function in `routes.py`

- Extract values from front-end user inputs
- One-hot encode categorical inputs (Store, StoreType, Assortment, StateHoliday)
- Build 1 observation to predict by appending all the feature values to *entered_li* in the exact same order as they are defined in the notebook.

Part 2.4: Heroku and submission

Deploy your final project using Heroku. Submit both your project and web app link on Canvas.

Extra Credit (2 pts each, total 6pts)

For every accomplished bullet points you can earn extra credit. This part is not required, but it is recommended to try working on at least one of these tasks:

1. Work on more advanced boosting models in Python notebook (e.g. XGBoost or LightGBM). Compare these models and elaborate on results.
2. Enable users to choose between different models, introducing drop-down menu with multiple options.
3. Next to the Sales prediction value, show model's accuracy to inform about the model's performance.