Take-Home Test

Thank you for considering the position at FlixBus. This take-home test is designed to gather more information about your experience and get a sense of how you approach problems we care about.

Problem Description

You are presented with a dataset `train_data.csv` that contains the daily sales for more than 1000 stores of a chain. In addition, we provide the dataset `store.csv` that contains details for each store.

Your Task

Your task is to build a model that predicts the daily sales for the test data set.

Please provide:

- Programming code, preferrably python for the soluton (and if not obvious -- instructions how to run it)
- Your reasoning about the problem and the solution you provided (data analysis, feature selection, model approach, ...)
- Your predictions in the form of a CSV file

For the first two parts you can pick your own format; it can be code + separate document, commented code, Jupyter notebook, etc.

Data details

Below we give you a brief description of the attributes of the CSV files:

train.csv and **test.csv**

- Id - an Id that represents a (Store, Date) duple within the test set
- Store - a unique Id for each store
- Sales - the turnover for any given day (prediction target)
- Customers - the number of customers on a given day
- Open - an indicator for whether the store was open: 0 = closed, 1 = open
- StateHoliday - indicates a state holiday. Normally all stores, with few exceptions, are closed on state holidays. Note that all schools are closed on public holidays and weekends.
    - a = public holiday, b = Easter holiday, c = Christmas, 0 = None
- SchoolHoliday - indicates if the (Store, Date) was affected by the closure of public schools

store.csv

- StoreType - differentiates between 4 different store models: a, b, c, d
- Assortment - describes an assortment level: a = basic, b = extra, c = extended
- CompetitionDistance - distance in meters to the nearest competitor store
- CompetitionOpenSince [Month/Year] - gives the approximate year and month of the time the nearest competitor was opened

- Promo - indicates whether a store is running a promo on that day
- Promo2 - is a continuing and consecutive promotion for some stores: 0 = store is not participating, 1 = store is participating
- Promo2Since [Year/Week] - describes the year and calendar week when the store started participating in Promo2
- PromoInterval - describes the consecutive intervals Promo2 is started, naming the months the promotion is started anew. E.g. "Feb,May,Aug,Nov" means each round starts in February, May, August, November of any given year for that store

Important note: You might wonder that certain information is missing in this take-home test or you might have questions. We expect you to make reasonable assumptions to resolve those and state these assumptions clearly so that we can evaluate how you deal with incomplete information and open questions.
Please be read to defend your assumptions and results, as well as explain how you came up with your solution!

If you should have questions concerning the task where you believe you cannot make assumptions, please feel free to reach out to team.aiiku@flixbus.com with your question.

Good luck and have fun!