PROJECT 3 RUBRIC: CHRISTOPHER STACK

Data Science Immersive | Project 3 | Instructor: John Marin
github link to project

Your project will be assessed using the following standards, derived from the data science workflow:

- [ ] Identify problem
- [ ] Acquire Data
- [ ] Parse Data
- [ ] Mine Data
- [ ] Refine Data
- [ ] Model Data
- [ ] Present Results

Acceptable performance for these standards is based on how well you've performed the specific requirements listed below.

SUMMARY:

Christopher: The stated goal was to find optimal location for a new liquor store in Iowa. The zip code offering new entrants the highest revenue per bottle (considering sales and margin) would best fit this criterion. This exercise, therefore, ultimately boiled down to data wrangling and the creation of pivot tables. The data cleaning included:

- 1) dollar signs from the dataframe
- 2) changed sale cost, bottle cost, bottle retail to float
- 3) added columns: price per liter = sale dollars/volume sold
- 4) fixed the zip code of Dunlap
- 5) converted the date to datetime type

Useful snippets of code related to the completion of this exercise included:

```
pd.pivot_table(df2,index=['County'],values = ['price_per_liter'])
pd.pivot_table(df2,index=['Zip Code'],values = ['price_per_liter'])
```

The regression models did not yield much actionable information (except that profit scaled with volume). Students who merged exterior demographic data related to latent variables (such as income from census

reports) received extra credit.

Performance Evaluation

> Instructors: Mark boxes with 'X' or "n/a" if a section does not apply. Note that "Exceeds Expectations" (3pts) is really only applicable for student writeups or specially marked "*Bonus*" Options.

Requirements	Incomplete (0)	Does Not Meet Expectations (1)	Meets Expectations (2)	Exceeds Expectations (3)
Identify: Write a detailed problem statement + Describe Goals & Criteria for Success	0			
Acquire: Import with Pandas & create at least one exploratory plot			2	n/a
Parse: State the assumptions of your data		1		n/a
Mine: Create necessary columns + Clean/format/process data			2.5	n/a
Refine: Determine outliers + distribution, correlation, & statistical analysis		1		*Bonus: handle outliers*
Model: Create at least one linear model; calculate probabilities + evaluate model, perform cross-validation, describe bias-variance tradeoff			2	*Challenging Bonus: perform Ridge and Lasso regularization and explain what they do*
Present: Create executive summary, present findings, summarize conclusions/recommendations, tailored for your audience (based on your scenario)		1		*Bonus: Include additional data & improvement suggestions*

Notes

Specific Positives:

- Nice cleaning of the data including the 712-2 zipcode issue
- Nice 'convert_money' function
- You used Adjusted R-Squared to gauge the value of your regressions (and you were correct that the regressions were essentially worthless)
- Accurate visual scatter plot in Cell 26
- Nice Regex and List Comprehensions to Pandas Datasets in Cell 62

Specific Negatives:

- Try not to regress on obviously correlated features such as

```
'year_sales ~ Q1_sales'
```

- Your cell 43 contained too many regressions at once that didn't add substantive value to your analysis
- You didn't have any clear goals written out at the outset of your report
- `df_graph = df.sample(frac=0.001)` --> Here a larger fraction would have been better for visualization purposes
- You didn't finish the assignment
- You left out the Data Refinement, Model Building, Plotting and Presentation of Results
- In Cell 7, you should have avoided a histogram of store*number, county*number, and item_number. These results are nonsensical.
- In Cell 8, you should have remarked the multicollinearity present between the volumetric measures of the alcohol in question as well as the bottles.
- Why did you leave 'shuffle false' in your KFold?
- You didn't provide very much written analysis of your conclusion or your milestone conclusions (although I know I saw a couple of sentences regarding the poorly performing regressions which I welcomed)

Score:

Based on the requirements, you can earn a maximum of **18** points on this project. Your score is: **9.5**

- Remember, your total score is helpful as a gauge of how well you met the project requirements, but use

the individual standards and instructor feedback to help iterate and improve on your projects!

PROGRESS REPORT

Student Check-in:

WHAT'S GOING WELL?	STRUGGLES	DEVELOPMENT PLAN
Data Wrangling and Cleaning, General Scaffolding of the Study	Completion, Written Analaysis, Articulation of Assumptions, Description of Results, Articulation and Focus on Goal	Use recently learned techniques to go back and complete this assignment so that it can figure into your portfolio of work. Please check in with instructor for specific guidance.