Hey Luna Lovegood

Whereever I've just given a grading, the text in the gradings sheet applies more or less. In case you have questions, slack me!

Flo

SQL

3/3

all (except question 9) answered

Python

Process Documentation

3/3

- great job on commenting your code properly!
- I especially like that you make good use of markdown, to explain your train of thought /reasoning and use comments to explain only code. You even chose to visualize information in tables! It's important to separate those two "tools" - markdown and code comments and you did great on that!

File Organization / GitHub Repo

3/3

- github structure neatly organized, everything is tidy. Great job!
- Keep in mind: Things like the `.ipynb_checkpoints` folder and `.DS_Store` (a MAC OS specific) file should be put into the `.gitignore` file (so that it is ignored by your git and not uploaded), because this is a file specific to your system. Check here for example.
- also: folder and file naming convention is to make everything lower case.
- The main README is of high quality. Very concise overview of your project. Awesome clickable TOC. Great pictures to illustrate everthing!
- In the **Objective** section you have a dead link to a picture and some markdown didn't work. But overall good job!
- I was especially amazed by the individual README files for the folders (!) and the apparent amount of work you guys put into those. Very impressive!

Code Readability

3/3

Nothing to criticize here. One can see that you're even making an effort to stick to the PEP8 style guide. Impressive! A few non-conformities here and there tough, but this comes with time.

Process - EDA, cleaning, pre-processing,

3/3

- I really like your ML flowchart picture in section 6.! Great jooob!

- the whole EDA/cleaning/preprocessing process was implemented, and you even iterated over these (trained & evaluated, came back to features, changed sth, trained & evaluated again, etc…) with smart reasoning to observe how your model quality changes. Great job.

- When it comes to plotting the correlation coefficient in a heatmap (multicolinearity), choose at best a diverging colormap (0 is special!) and use `center=0`

- Great job regarding the preprocessing! And you did a wonderful job with training a model on "fairly" unprocessed data. But a bit of a nitpicking point: creating new columns for `month` and `quarter` would already be a tiny bit of feature engineering which would be put after the first model iteration… but as said… this is just nitpicking ;.)

- Here and there you could have added some more observation instead of *just* plotting. Like: *As we can see from these barplots, the variables* `sqft_above`, … *show a skewed normal distribution. Variable* `sqft_lot15` *features a long tail…*

- of course the plots in these notebooks serve as tools for us data analysts, and therefore things like a grid could be useful. But keep in mind to reduce unnecessary plot elements (a grid is not necessary oftentimes) when you create plots for communicating insights with them.

Number of models compared

3/3

- great job on the number of models being compared!
- even for the first rough model, a `y_pred` vs `y_test` plot would have been nice.
- impressive score and improvement for the second model after adding the OHE'd categoricals and scaled/normalized features!
- You could have added some interpretation for your `y_pred` vs `y_test` plot of the KNN model. Here you see that the model performs better for lower housing prices and increasingly worse towards higher housing prices (the dots fan out! also called "heteroscedasticity" - terrible word). KNN probably performs poorly because towards higher house prices, the datapoint density decreases and towards higher prices there are simply less and less "neighbors" found when the algorithm is trained.

Tableau

Dashboard Clarity

3/3

KPIs

3/3
You really excel in highlighting the key numbers here (just one example: That slide with the picture of the Seattle neighborhood - fantastic!)

Clear and informative visual design

3/3

- the design is very clear and informative. Sizes of fonts and plot dimensions is very readable etc.
- Regarding your Tableau story, in particular the scatterplots. I see you made an effort to tackle "overplotting" a bit (probably by lowering the alpha level. But you could take that to a next level, e.g. make the markersize a bit smaller or opt for a density plot. You can check our Additional Resources on Notion for techniques on how to avoid overplotting. One other tiny issue: For me it wasn't imediatly be unclear to me what the colorcode of the dots means.
- Great and clean line plots with informative text fields! Those really help to contribute to the "story"
- For your distribution plots ("A. Distribution…."): If you decide to give an increasing ordinal variable (number of bedrooms) different colors, try a colormap that is perceptually uniform and sequential (cool to warm, bright to dark etc… in matplotlib e.g. viridis, plasma, …)

All metrics clearly identified and plotted

Answer the challenge questions

3/3
all questions answered.

Presentation

Time

2/3
7:00 minutes, a bit overtime

Storytelling

2/3

- Quite well structured!
- Well crafted result slide. The x- and y-labels (y-predicted vs. y-test) could have been bigger though and labeled for a non technical audience like "predicted price in $" and "true price in $". Imagine you present this to a manager or another department not proficient or familiar with pandas/python/sklearn etc.
- at slide INSIGHTS you had technical difficulties and you, Marian took over. Excellent improvisation - bloopers happen! When Antonios mic worked again he (accidentally?) interrupted the flow. In my point of view, Marian should have cracked on and Antonio should have taken over only at the next slide. The way it happened, to me it sounded like you Antonio explained Marian's points again, no?

Visual Desing: visually appealing

3/3

- stunning visuals, and consistent design! Here is indeed where you guys shine.
- great making use of visual help (logos, icons etc.). The slide "Processes" is a prime example for how you can craft a slide perfectly for a target audience. We all know the logos and symbols by now, so there's no need in writing text. Great job!

Text + image balance

3/3

- excellent text to image balance!
- very good speech to visual information ratio!

Use of vocabulary

3/3