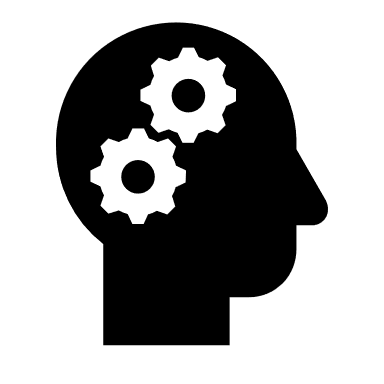
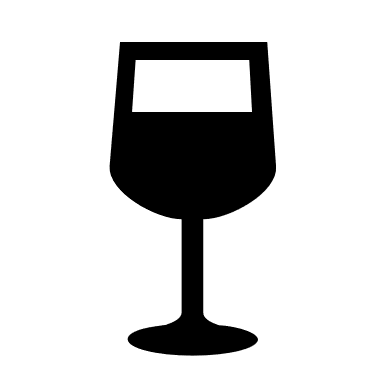
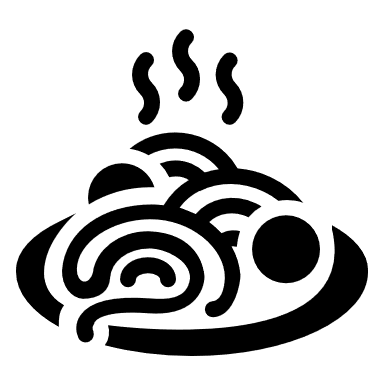
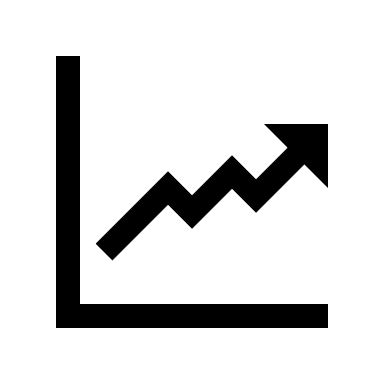
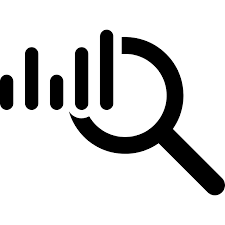
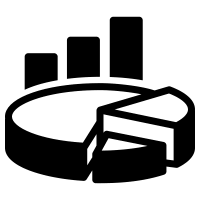
***********Le Château du Technologie Permettant *** **

**~ General Analytics ~ **

*The following include complimentary side-order of reporting. Choose from a convenient browser-readable readable “notebook” or PDF report via e-mail. Ask your server about upgrading to in-person and Power-Point options. Popular analytics tasks might be eligible for upgrade into a dashboard or pipeline.*

Data Cleaning & Restructuring $*Easy to Complex*

We can help “clean” non-standardized data that by reformatting, removing stray characters, etc. An example might be getting rid of redundant categories by turning “True”, “true”, “T”, “false”, “F”, and “False” into “True” and “False”. This may also involve transforming (‘transposing’) rows into columns or simple “feature engineering” (creating columns by extrapolating from existing ones).

* *The price ranges from easy to complex depending on how “messy” the data is. Restructuring data generally increases the work involved.*

Data Exploration $*Straightforward to Involved*

We can review datasets to get summative stats about them, produce basic visualizations, and report back any odd findings that pop out. This is generally just to get an idea of what is going on, to make sure we aren’t missing anything obvious, and to formulate questions for further analysis.

* *The price ranges from straightforward to fairly involved depending on the nature of the dataset.*

Specialty!: Research Question $*More Involvement & Effort*

Basically, an exploration with a specific question in mind – “How often do we have to remanufacture products, and what factors feed into that?”, “Does there seem to be a correlation between variable X and variable Y?”. An example order can be found in [gunsFebInvest.pdf](http://nbviewer.jupyter.org/github/sweeney-th/Portfolio/blob/master/DataScience/GunDeathsInFebInvestigation/gunsFebInvest.pdf).

* *This dish often involves at least a little a little bit of the previous two items, so it is generally safe to assume it takes more effort.*

**~ Dashboards ~** 

*Dashboards are a great way to visualize reusable, routine reports in a browser, many of which are interactive. They can be constructed to read in data files and create interactive visuals based on them. There two main ways to cook the up, ask your server which is right for your project.*

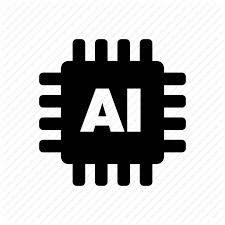
Proprietary Analysis Tools $*Varies*

Dashboards can be created using software like PowerBI and Spotfire. These are made using drag-and-drop interfaces to create custom Dashboards intuitively. This makes them easier to use, but less flexible.

Custom Dashboards $ *Varies on Flexibility need*

If tools like Spotfire and PowerBI are insufficiently flexible, a custom dashboard for which the server writes the underlying analysis might be possible.

* *The price of customizable dashboards also varies but is generally higher than that of Proprietary Analysis Tools to account for increase in flexibility.*

**~ Machine Learning ~ **

*Machine Learning (ML) tasks generally require larger datasets (thousands of rows) but may be able to produce classifiers, clustering, or regressions models to help predict or understand datasets. They are not right for every project but can be extremely powerful if used appropriately. Broadly speaking, they work by ‘learning’ from ‘training’ sets fed to them by the programmer and then applying the patterns they observe to new information or by grouping observations into “clusters” of similar data. Ask your server for a consultation. Machine Learning projects are served with a presentation explaining the underlying algorithms and modeling techniques. A full example of a Machine Learning project can be seen in* [*signatureProject.pdf*](http://nbviewer.jupyter.org/github/sweeney-th/Portfolio/blob/master/DataScience/TwitterClassifier/signatureProject.pdf)*.*

* General ML Pricing: *ML projects should generally be given more time than non-ML projects. Depending on the complexity and size of the data, these dishes may take longer to prepare and serve. However, with clean data, results might be seen surprisingly fast.*

Classification $*Simple to Involved*

Classifiers attempt to place observations into categories. They are available in binomial (two categories like “malignant” or “benign”) or multinomial (>2 categories) varieties. Some classifiers can produce “rules” like a flowchart that a person can easily understand and apply to new cases. See [spam.pdf](http://nbviewer.jupyter.org/github/sweeney-th/Portfolio/blob/master/DataScience/Spam/spam.pdf) for an example of a binomial classifier.

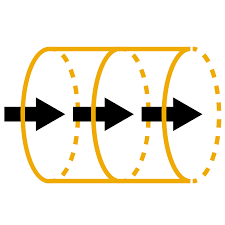
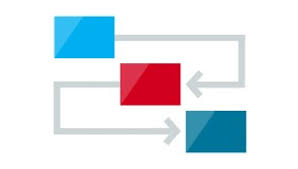
* *The price of classification increases with the number of categories. Multinomial classifiers are generally more involved.*

Regression $*Mostly Simple*

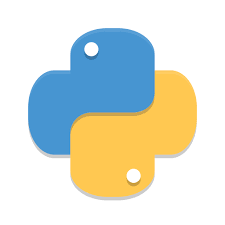
Regression models generally work on “continuous” variables within a range, like the price of a house, as opposed to True/False values. Broadly speaking, they work by creating a formula for a line that describes the data (think *y = mx + b*) and applying it to new information.

Clustering $ *Simple to Involved*

Clustering models don’t predict anything based on previous data in the way that classifier or regression model might. Instead they generate the categories by grouping similar observations together. A clustering that analyzes social media text, for instance, might sort people into categories which a human interpreter would know as “liberal”, “conservative”, or “centrists”.

**~ Programming & Pipelines~ **

*Custom programs to complete desired task can be requested and written from scratch. If desired, several programs can be strung together into a “pipeline” that can be run periodically to automate certain analysis. This is the most varied category so ask your server what is in season. An example of a pipeline can be found in* [*bioinformatic\_functions.ipynb*](https://github.com/sweeney-th/Portfolio/blob/master/DataScience/Bioinformatics/bioinformatic_functions.ipynb)*.*

**~ Tech Education ~ **

*We are happy to provide explanations of how ours tools work to curious folks, give people some guidance, or even offer some teaching if you’re looking to skill up! Generally, we are proficient in Python and R, but ask your server what other areas they can teach.*