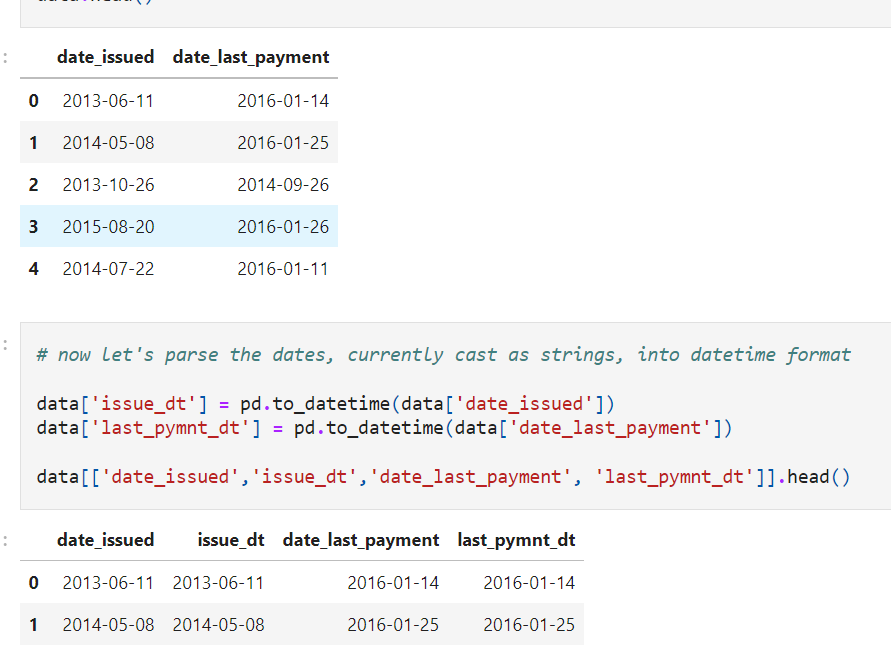
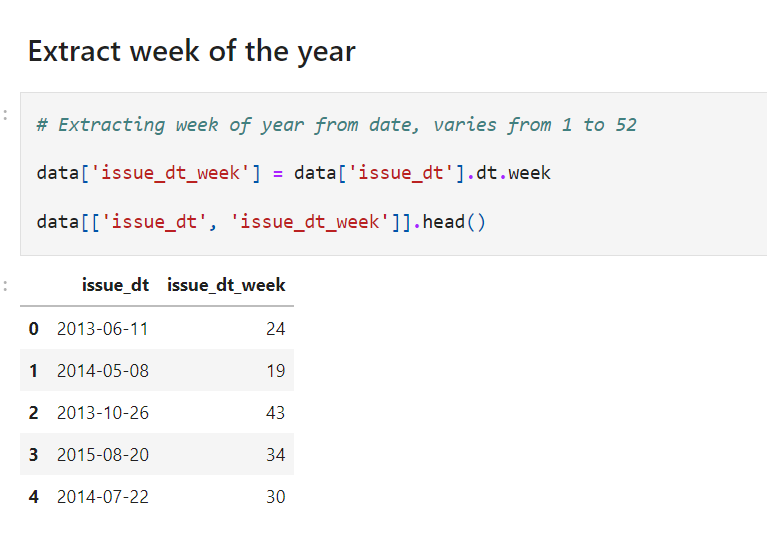
**Engineering Dates**

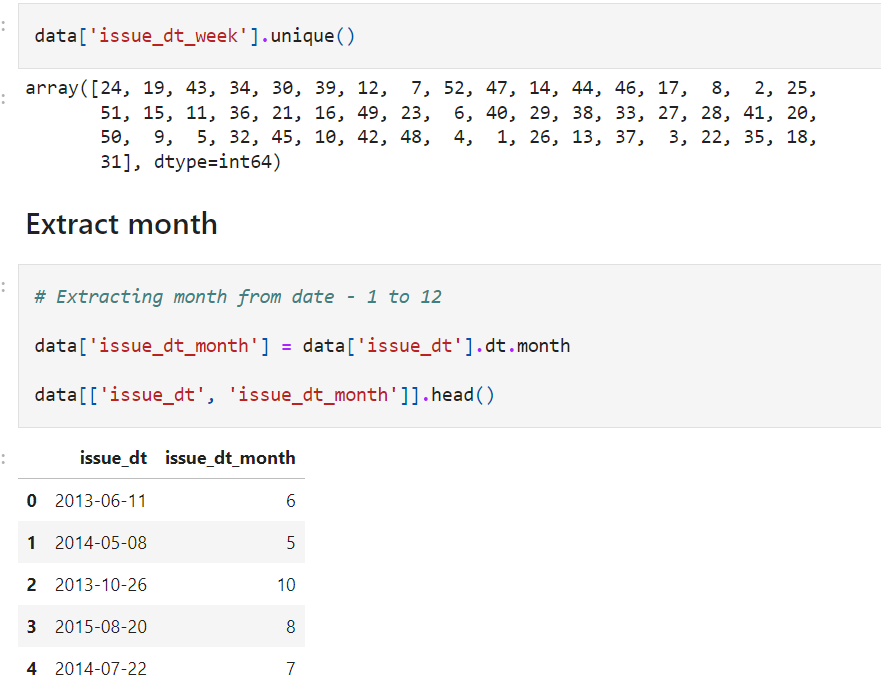
Date variables are special type of categorical variable. By their own nature, date variables will contain a multitude of different labels, each one corresponding to a specific date and sometimes time. Date variables, when preprocessed properly can highly enrich a dataset. For example, from a date variable we can extract:

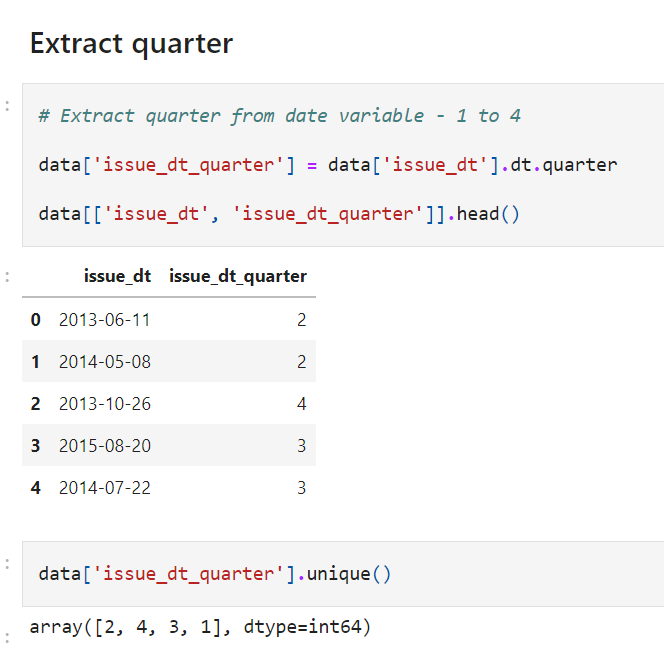
* Week of the year
* Month
* Quarter
* Semester
* Year
* Day (number)
* Day of the week
* Is Weekend?
* Time differences in years, months, days, hrs, etc.

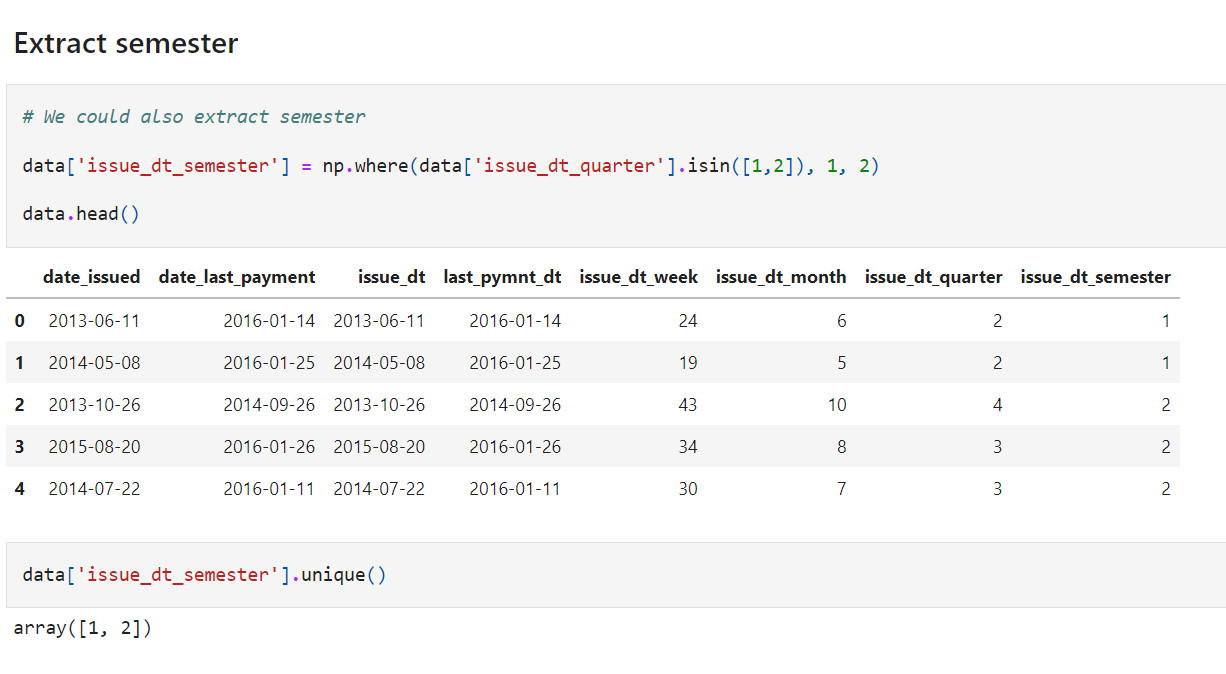
Date variables should not be used as categorical variables when building a machine learning model. Not only because they have a multitude of categories, but also because when we actually use the model to score a new observation, this observation will most likely be in the future, an therefore its date label, might be different from the ones contained in the training set and therefore the ones used to train the machine learning algorithm.



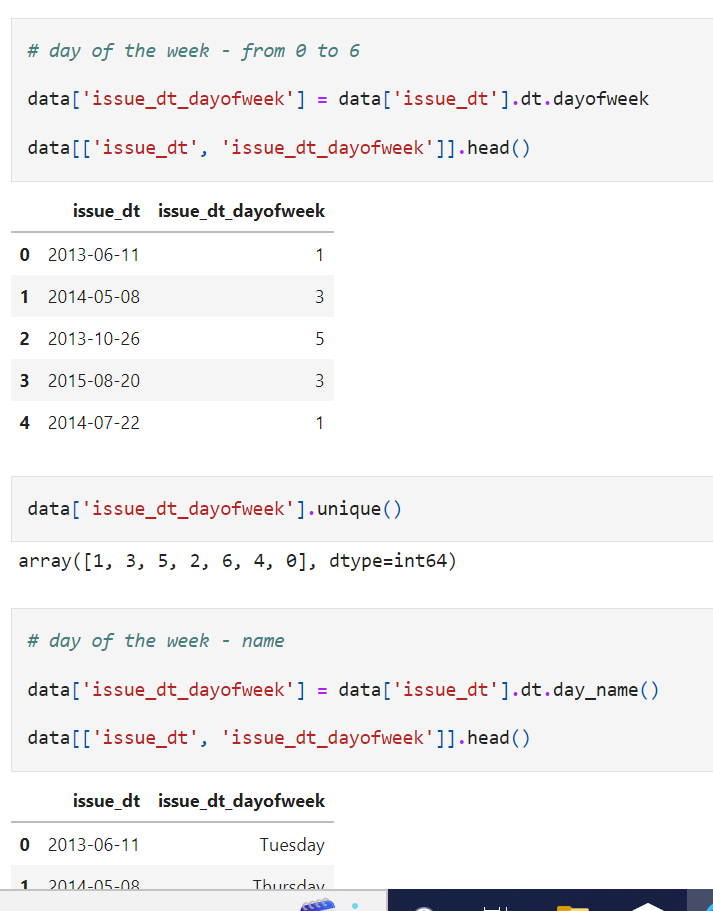


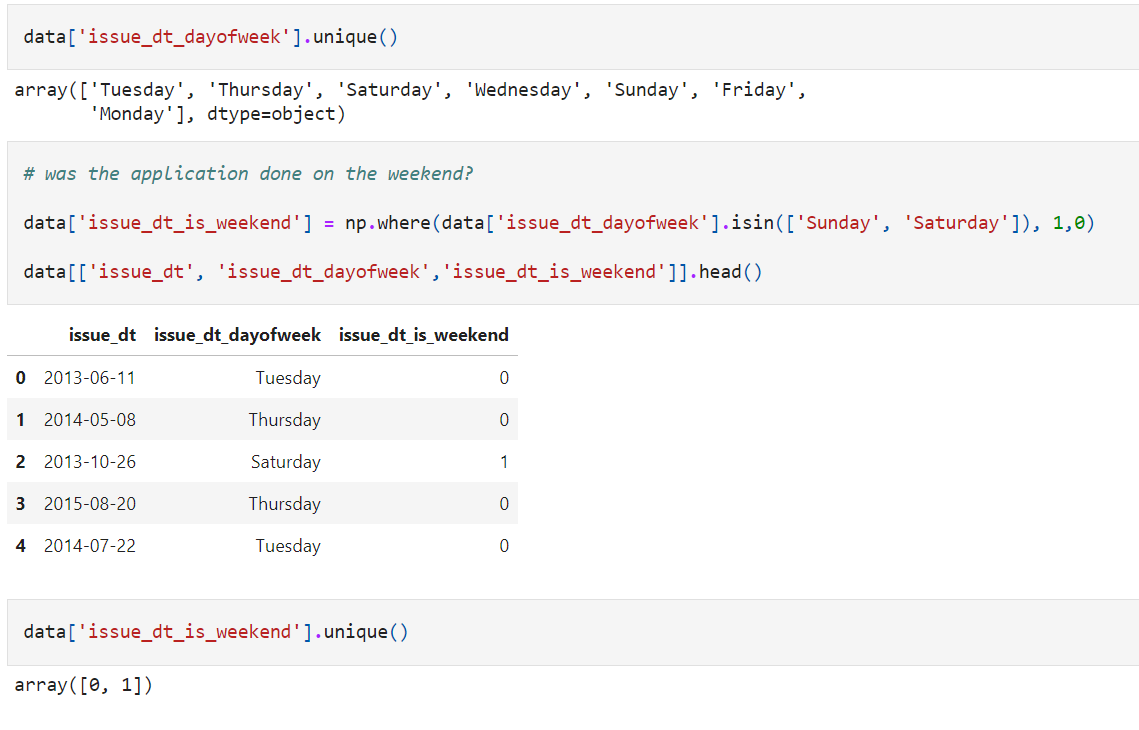




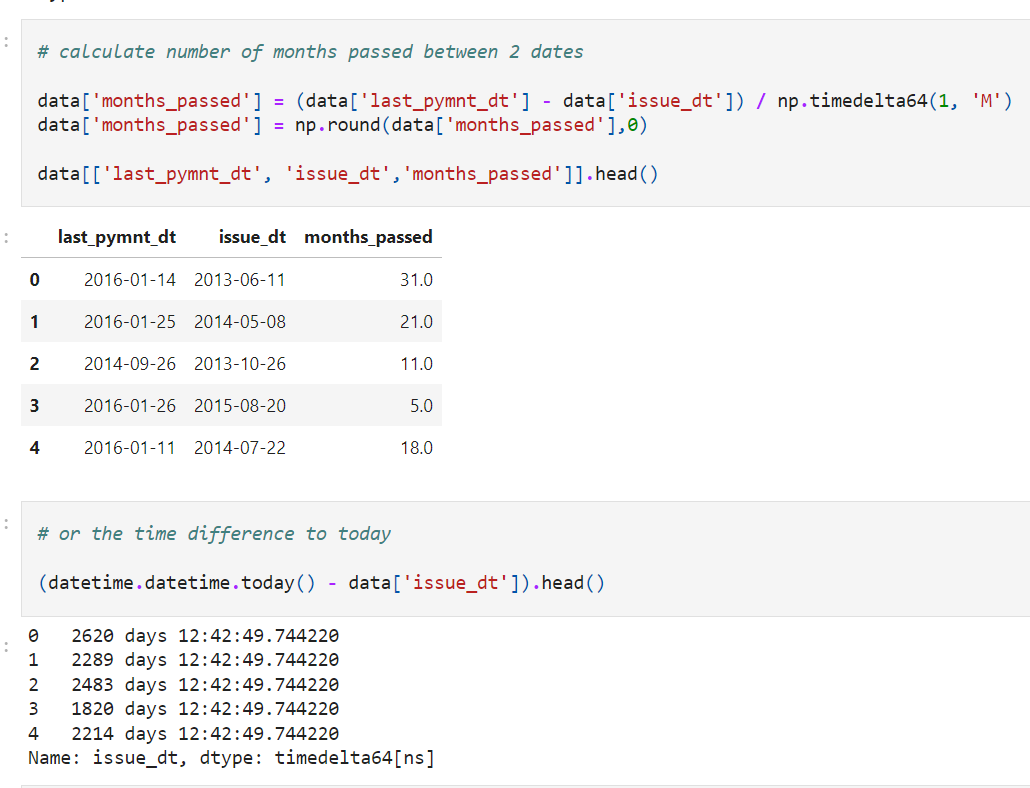










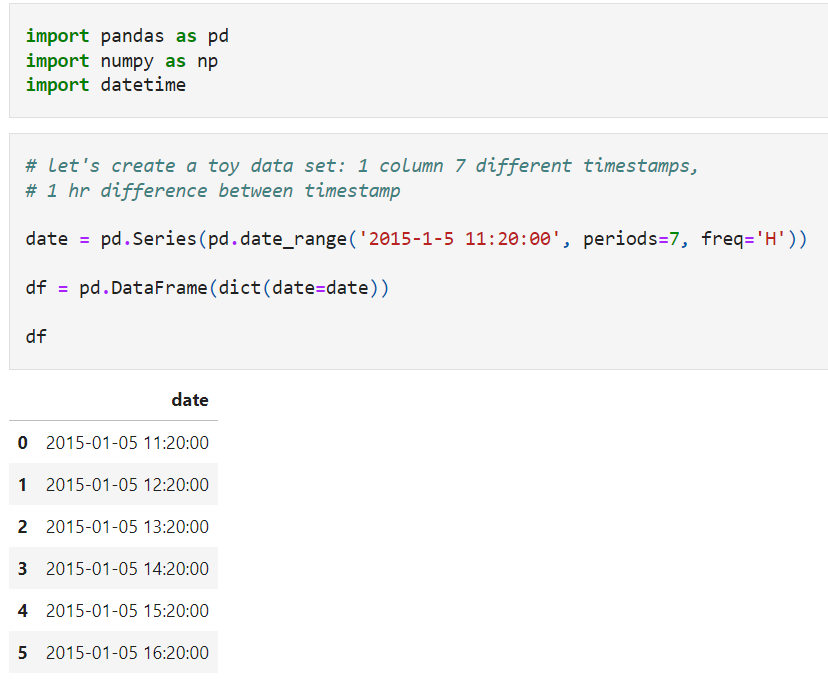


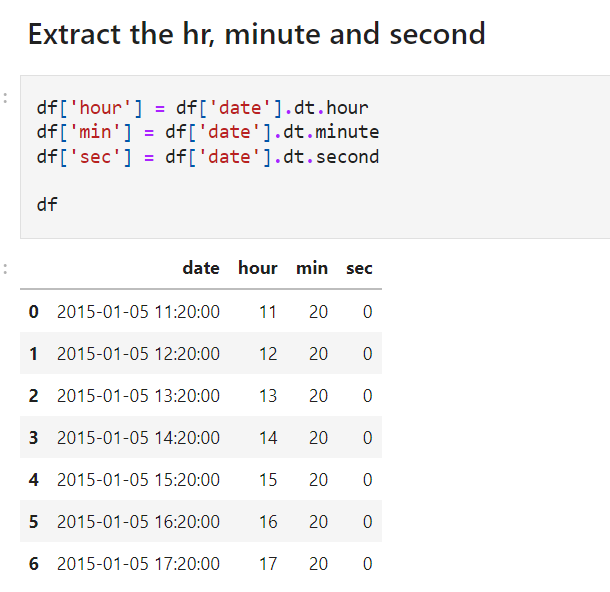
**Engineering Time**

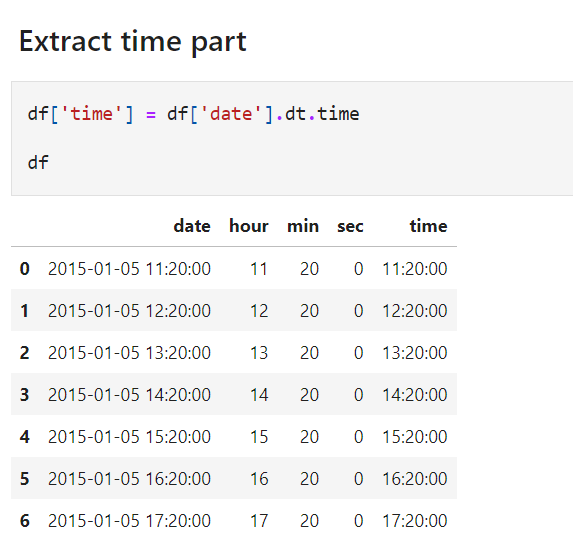
In this demo, we are going to extract different ways of representing time from a timestamp. We can extract for example:

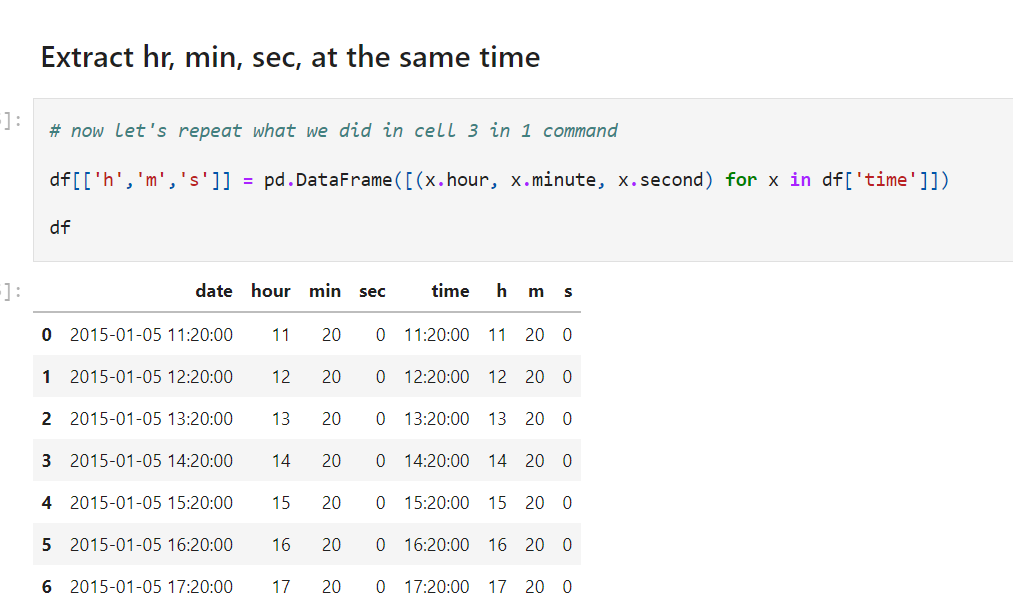
* hour
* minute
* second
* data
* elapsed time

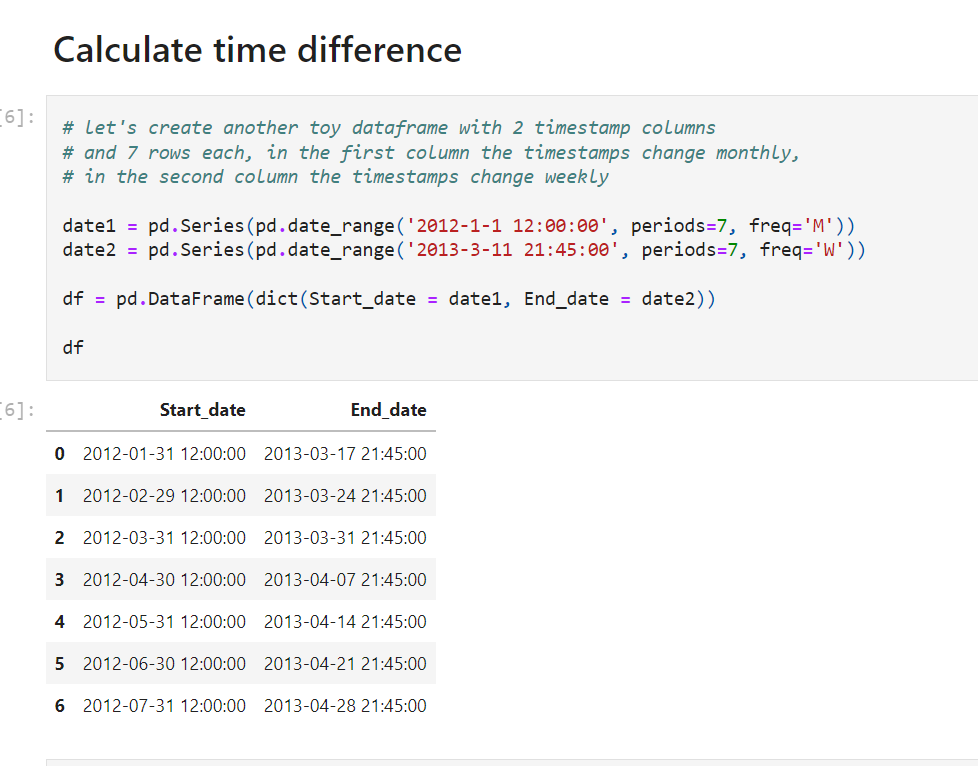
We will create a toy dataset for the demonstration.

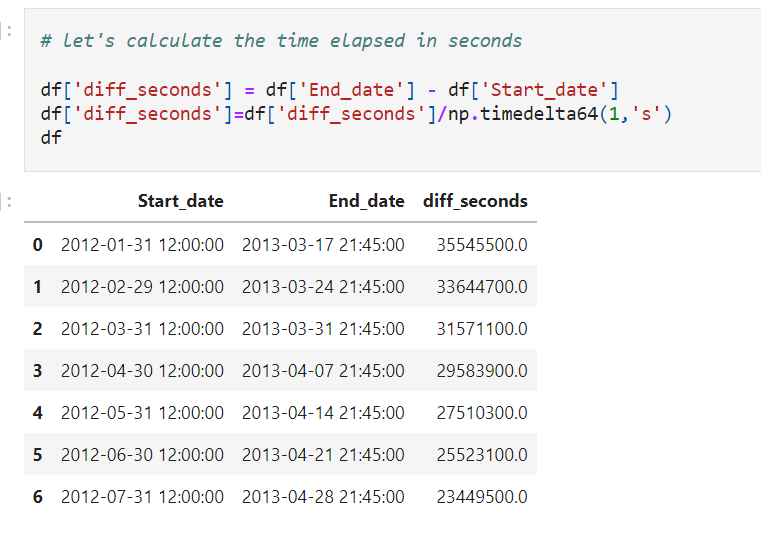


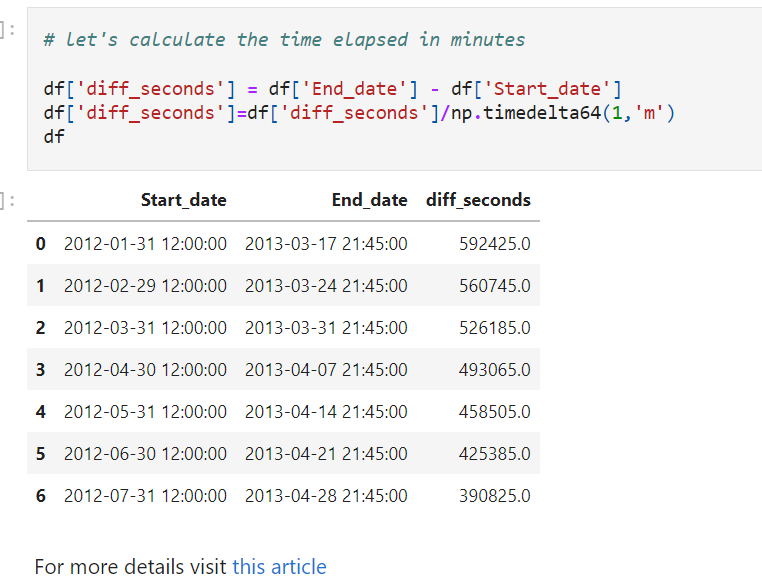


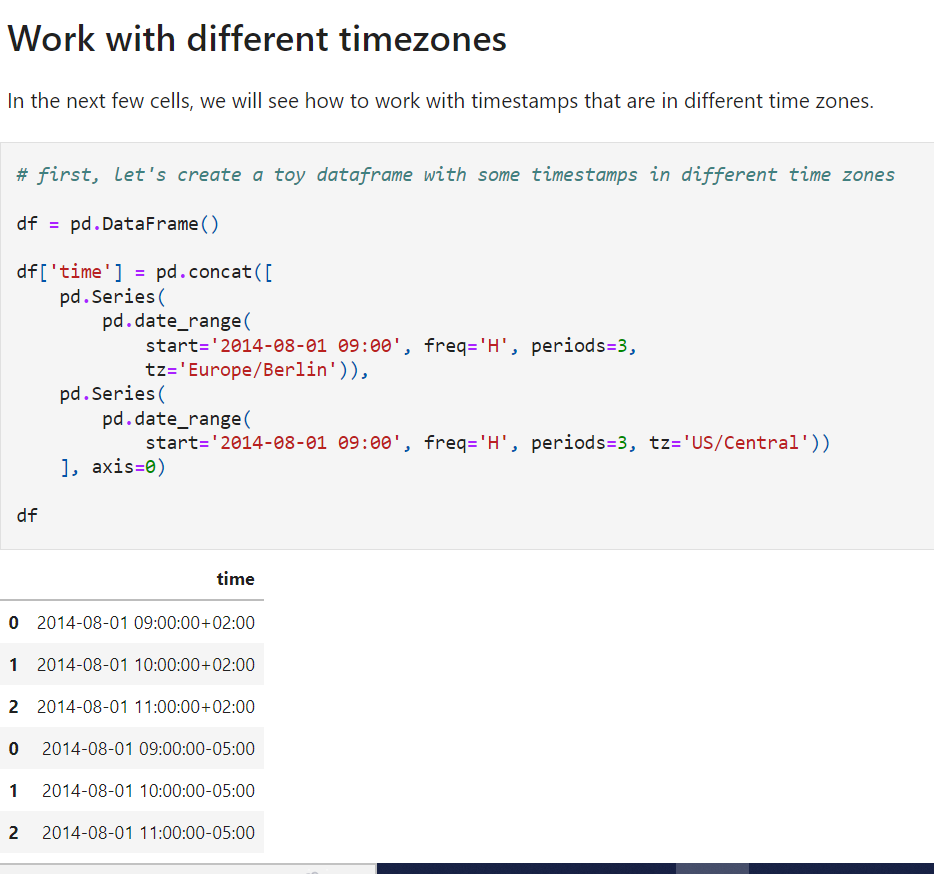


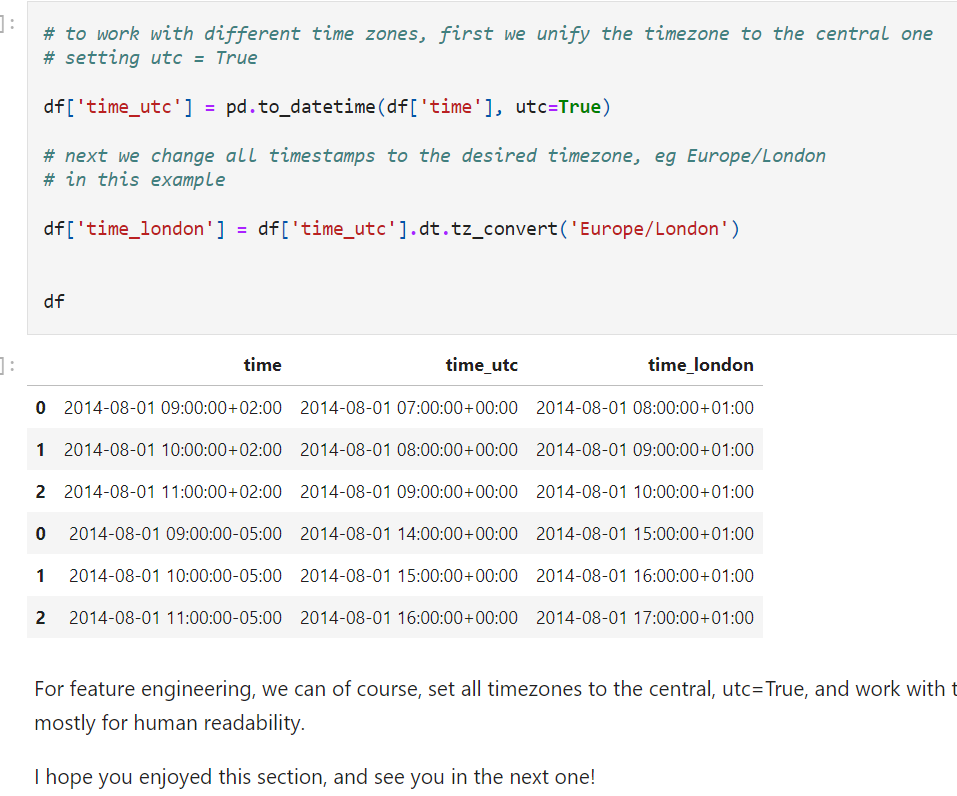












PIPELINE

[feature-engineering-for-machine-learning/Section-13-Putting-it-altogether at main · solegalli/feature-engineering-for-machine-learning (github.com)](https://github.com/solegalli/feature-engineering-for-machine-learning/tree/main/Section-13-Putting-it-altogether)

The challenge:

Beat the performance of my Lasso regression by using different feature engineering steps **ONLY!!**.

The performance of my current model, as shown in this notebook is:

1. test mse: 1063016789.3316755
2. test rmse: 32603.938248801718
3. test r2: 0.8453144708738004

To beat my model you will need a test r2 bigger than 0.85 and a rmse smaller than 32603.

================================================================================================

Conditions:

* You MUST NOT change the hyperparameters of the Lasso.
* You MUST use the same seeds in Lasso and train\_test\_split as I show in this notebook (random\_state)
* You MUST use all the features of the dataset (except Id) - you MUST NOT select features

================================================================================================

If you beat my model:

Make a pull request with your notebook to this github repo: <https://github.com/solegalli/udemy-feml-challenge>

And add your notebook to the folder:

-StudentsSolutions\_v1.0.0

How to make the PR

1. fork the repo:

Go to <https://github.com/solegalli/udemy-feml-challenge>, and click on the **fork** button at the top-right

1. clone your forked repo into your local computer:

* Go to [www.github.com/yourusername/udemy-feml-challenge](http://www.github.com/yourusername/udemy-feml-challenge)
* Click the green button that says clone or download
* copy the url that opens up
* power up a git console
* type: git clone (paste the url you copied from github)
* done

1. Make a copy of the jupyter notebook and add your name:

* Open up the Jupyter notebook called 13-Assignement.ipynb
* Click the "File" button at the top-right and then click "Make a copy"
* **Work your solution in the Copy** and not in the original assignment (otherwise there will be conflicts when making the PR)
* Change the name of the copy of the notebook to: 13-Assignement\_yourname.ipynb
* Move the notebook to the folder **StudentsSolutions\_v1.0.0**
* done

When you finish, just commit the new notebook to your fork and then make a PR to my repo.

* git add StudentsSolutions\_v1.0.0/13-Assignement\_yourname.ipynb
* git commit -m "your commit message"
* git push origin master or git push origin yourfeaturebranch
* go to your repo and make a pull request.

But i have a notebook ready and I haven't cloned the repo yet, how can I make the PR?

If you worked in the copy you downloaded from Udemy before forking and cloning this repo, then follow this steps:

1. fork the repo:

Go to <https://github.com/solegalli/udemy-feml-challenge>, and click on the fork button at the top-right

1. clone your forked repo into your local computer:

Go to [www.github.com/yourusername/udemy-feml-challenge](http://www.github.com/yourusername/udemy-feml-challenge)

* Click the green button that says clone or download
* Copy the url that opens up
* Power up a git console
* Type: git clone (paste the url you copied from github)
* Done

1. Rename your solution as follows and copy it into your cloned repo:

* Rename your solution notebook to: 13-Assignement\_yourname.ipynb
* Copy this file into the cloned repo, inside the folder **StudentsSolutions\_v1.0.0**
* Done

When you finish, just commit the new notebook to your fork and then make a PR to my repo

* git add StudentsSolutions\_v1.0.0/13-Assignement\_yourname.ipynb
* git commit -m "your commit message"
* git push origin master or git push origin yourfeaturebranch
* go to your repo and make a pull request.

**Good luck!!**

**Questions for this assignment**

How else could you impute  the  missing data?

How  else could you encode the categorical variables?

How could you trasform the numerical variables?

Could you create new features from the existing ones?