# Data Science final project

Many apologies for the late submission – I have been little under the weather since Thursday but feeling a lot better now. I don’t think it was Corona but who knows!

Regarding my final project, I have left things a little late due to spending more time on the theory and practises, and will spend as much time as possible till next lesson on exploring the data sets.

My first choice I believe may be hindered by the dataset itself (I have provided links below).   
So I have listed two other ideas for final project and will book time with Oliver as soon as possible so get the help I need.

## 1st choice

|  |  |
| --- | --- |
| **Project Proposal:** | Analyse trends of live births in England and Wales, based on country of birth of Father and country of birth of Mother, with goal to predict percentage of children born to parents born outside of UK in 2030 (with potential uses in determining most risky time periods for changes in immigration policies) |
| **Project Brief:** | Only datasets available are from [Office of National Statistics](https://www.nomisweb.co.uk/query/select/getdatasetbytheme.asp?opt=3&theme=&subgrp=) (“Live births in England and Wales by characteristics of mother and father”).  The largest/smallest groups of live births regarding mothers’ country of birth  The largest/smallest groups of live births regarding fathers’ country of birth  Births based on mothers’ country of birth visualized  Average age of mother and father  Whether child is first-born, second-born etc. and which is more common and least common  Age differences between first-born and second-born and third-born (and average)  Percentage of live births to mother born in UK  Percentage of live births to mother born outside of UK  Percentage of live births to father born outside of UK  **RISK** – the dataset is not large enough? |
| **Technical Env:** | <http://localhost:8891/notebooks/Untitled.ipynb?kernel_name=python3>  Set up local git repository for project, connected to public remote GitHub repository: |

## 2nd choice

|  |  |
| --- | --- |
| **Project Proposal:** | Analyse trends of global shark attacks, from 1901 to the present, with goal to predict most likely area to suffer from aggressive shark attacks. |
| **Project Brief:** | Dataset available from Kaggle:  <https://www.kaggle.com/teajay/global-shark-attacks>  <http://www.sharkattackfile.net/index.htm>  Which species is the greatest aggressor?  What injuries prove most fatal?  Attack types (visualized)  Country of attacks (visualized)  What activities prove most risky?  Any correlations between activity and fatal attacks?  Any correlations between time and fatal attacks?  Average age of victims  Average time of attacks  (min and max values for year needed – is there enough data?) |
| **Technical Env:** | <http://localhost:8891/notebooks/Untitled.ipynb?kernel_name=python3>  Set up local git repository for project, connected to public remote GitHub repository: |

## 3rd choice - taken as example of project from DataCamp

|  |  |
| --- | --- |
| **Project Proposal:** | Analyse any trends of Nobel Prize Winners, from 1901 to the present, with goal to predict category of winner in 2021. |
| **Project Brief:** | Dataset available from Kaggle: <https://www.kaggle.com/deepakdeepu8978/nobel-laureates-1901present>  Who gets the Nobel Prize?  Country dominance (visualized)  What is the gender of a typical Nobel Prize winner?  The first woman to win the Nobel Prize  Any repeat winners?  What is average age of winner?  Age differences between prize categories  Oldest and youngest winners |
| **Technical Env:** | <http://localhost:8891/notebooks/Untitled.ipynb?kernel_name=python3>  Set up local git repository for project, connected to public remote GitHub repository: |