
PREDICTIVE ANALYTICS

Assignment Project Exam Help

INDIVIDUAL COURSEWORK

<https://powcoder.com>

Add WeChat powcoder

A P MOORE

- Friday 26th February 2021
- 60% of module mark
- 2000 words

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

The individual coursework task is to identify a dataset and explore building a predictive model using the methods and techniques presented in the first 5 weeks of the course.

Assignment Project Exam Help

There are six main steps:

1. Obtain a dataset and explain the problem you are trying to solve.
This will characterise the type of predictive model you can build
2. Explore the data to gain insights.
Visualize and explain the main trends in the data
3. Prepare the data to better expose the underlying data patterns to Machine Learning algorithms.
4. Explore different models and shortlist the best ones.
5. Fine-tune your models and combine them into a better solution.
6. Present your final solution with any summary conclusions.

Appendix B. Machine Learning Project Checklist

This checklist can guide you through your Machine Learning projects. There are eight main steps:

1. Frame the problem and look at the big picture.
2. Get the data.
3. Explore the data to gain insights.
4. Prepare the data to better expose the underlying data patterns to Machine Learning algorithms.
5. Explore many different models and shortlist the best ones.
6. Fine-tune your models and combine them into a great solution.
7. Present your solution.
8. Launch, monitor, and maintain your system.

Obviously, you should feel free to adapt this checklist to your needs.

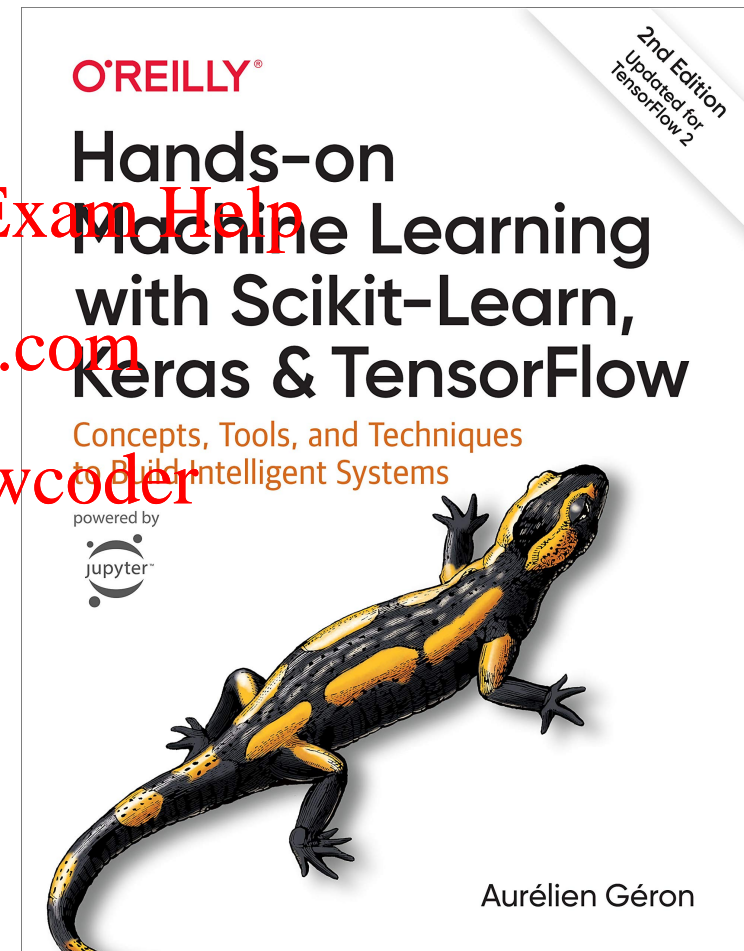
Frame the Problem and Look at the Big Picture

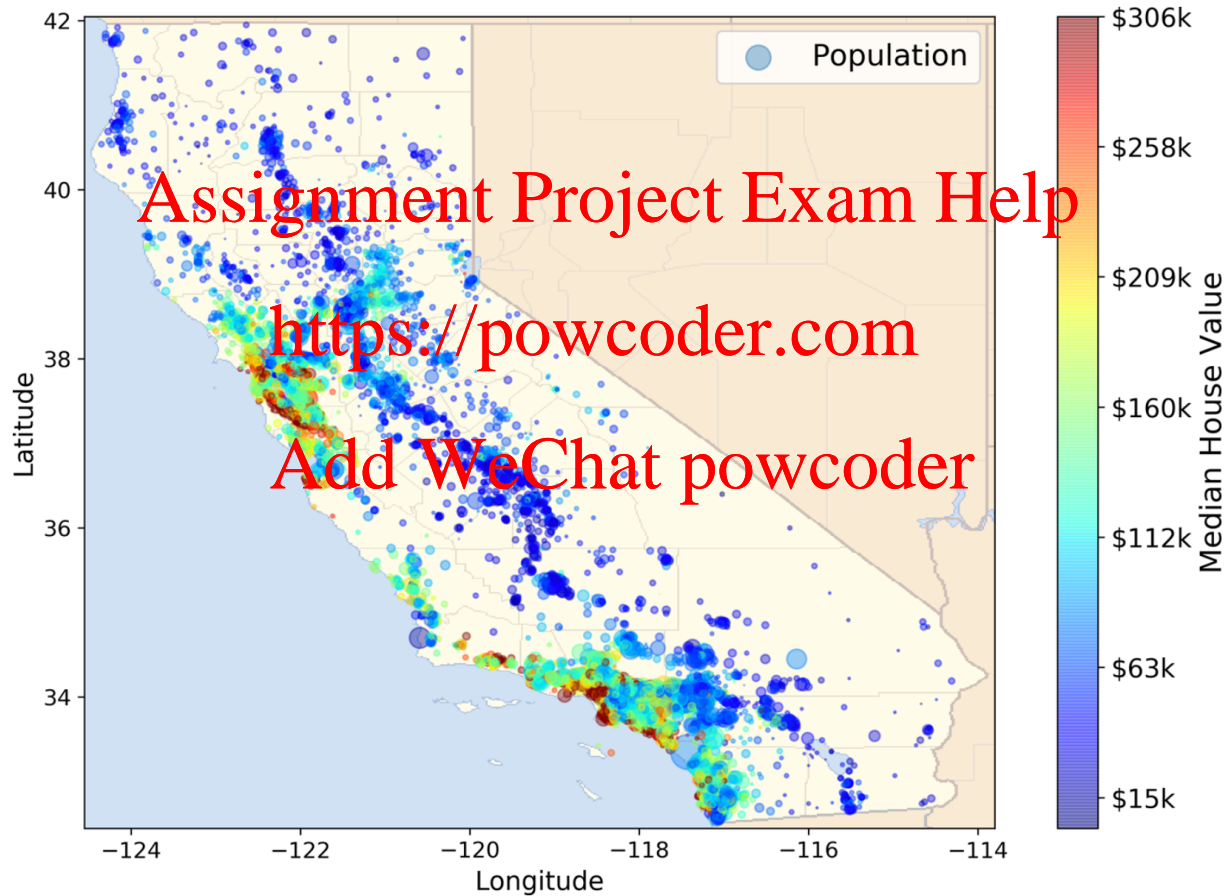
1. Define the objective in business terms.
2. How will your solution be used?
3. What are the current solutions/workarounds (if any)?

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder





```
In [7]: housing["ocean_proximity"].value_counts()
```

```
Out[7]: <1H OCEAN    9136
        INLAND    6551
        NEAR OCEAN 2658
        NEAR BAY   2290
        ISLAND      5
        Name: ocean_proximity, dtype: int64
```

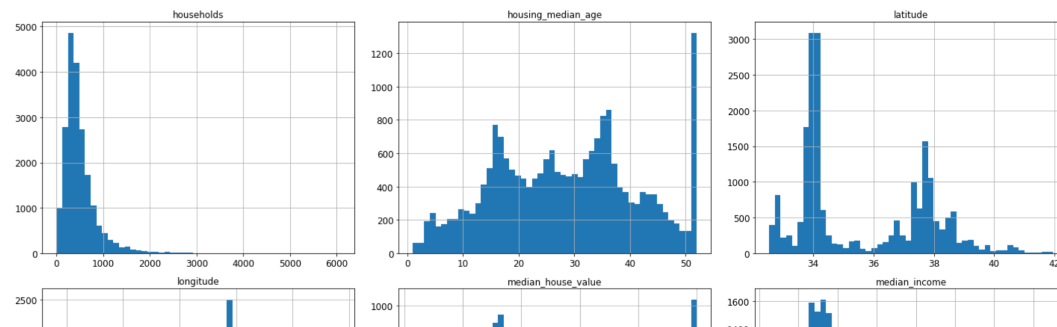
```
In [8]: housing.describe()
```

```
Out[8]:
```

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value
count	20640.000000	20640.000000	20640.000000	20640.000000	20640.000000	20640.000000	20640.000000	20640.000000	20640.000000
mean	-119.569704	35.631861	28.639486	2635.763081	537.870553	1425.476744	499.539680	3.870671	206855.816909
std	2.003532	2.135952	12.585558	2181.615252	421.385070	1132.462122	382.329753	1.899822	115395.615874
min	-124.350000	32.540000	1.000000	2.000000	1.000000	3.000000	1.000000	0.499900	14999.000000
25%	-121.800000	33.130000	18.000000	1447.500000	206.000000	187.000000	240.000000	2.563400	119600.000000
50%	-118.490000	34.260000	29.000000	2127.000000	435.000000	1166.000000	409.000000	3.534800	179700.000000
75%	-118.010000	37.710000	37.000000	3148.000000	647.000000	1725.000000	605.000000	4.743250	264725.000000
max	-114.310000	41.950000	52.000000	39320.000000	6445.000000	35682.000000	6082.000000	15.000100	500001.000000

```
In [9]: %matplotlib inline
import matplotlib.pyplot as plt
housing.hist(bins=50, figsize=(20,15))
save_fig("attribute_histogram_plots")
plt.show()
```

Saving figure attribute_histogram_plots



Useful places for ML datasets:



- Tabular & cleaned: <https://github.com/EpistasisLab/pmlb/tree/master/datasets>
- By domain: <https://datasetlist.com>
- By application: <https://github.com/awesomedata/awesome-public-datasets>
- Search engine: <https://datasetsearch.research.google.com>

Assignment Project Exam Help

https://powcoder.com

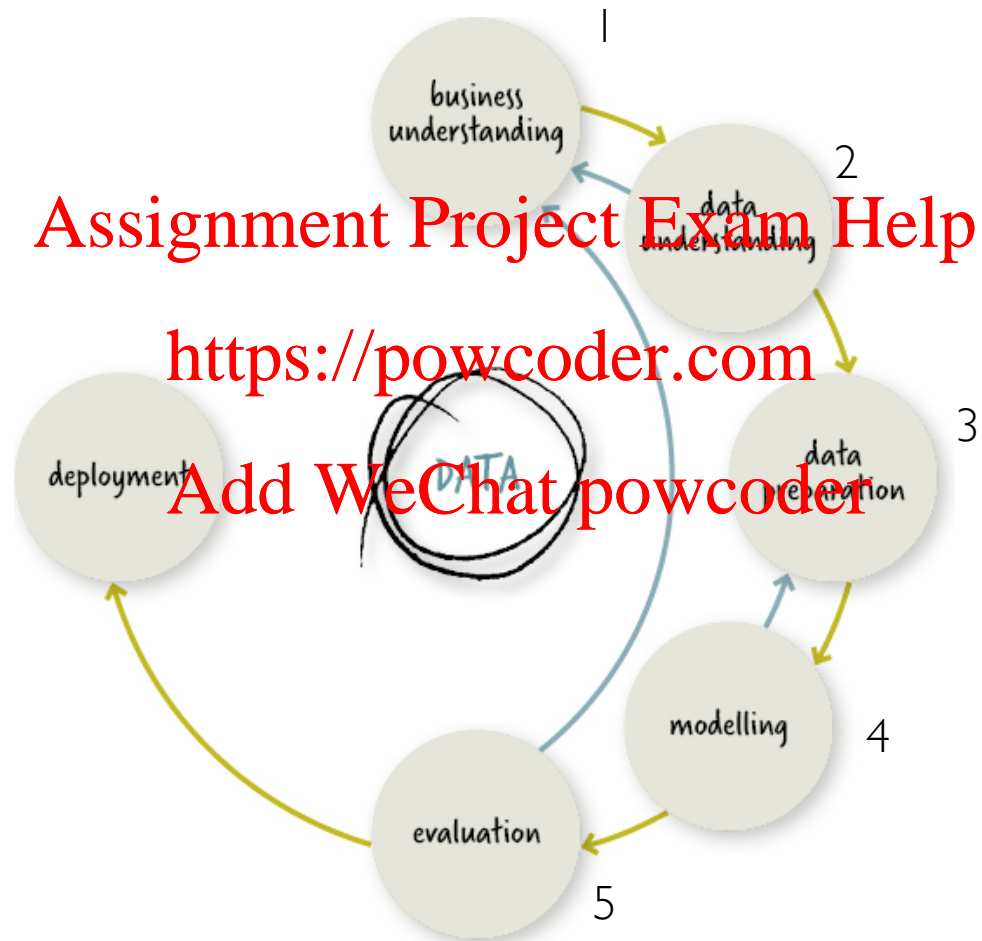
Add WeChat powcoder

Economics

-  American Economic Association (AEA)
-  EconData from UMD
-  Economic Freedom of the World Data
-  Historical Macroeconomic Statistics
-  INFORUM - Interindustry Forecasting at the University of Maryland
-  DBnomics – the world's economic database - Aggregates hundreds of [...]
-  International Trade Statistics
-  Internet Product Code Database [fixme]

CRISP CYCLE

DATA DEVELOPMENT LIFECYCLES



Fast First Pass

Make a first-pass through the project steps as fast as possible. This will give you confidence that you have all the parts that you need and a baseline from which to improve.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Cycles - The process is not linear but cyclic. You will loop between steps, and probably spend most of your time in tight loops between steps 3-4 or 3-4-5 until you achieve a level of accuracy that is sufficient or you run out of time.

The write up in the final submitted Notebook can be more linear - you do not need to include all of your work, ie. including all dead-ends, and it should be concise and consistent.

Attempt Every Step

It is easy to skip steps, especially if you are not confident or familiar with the tasks of that step. Try and do something at each step in the process even if it does not improve accuracy. You can always build upon it later. Don't skip steps, just reduce their contribution to your final submission as necessary.

Assignment Project Exam Help

<https://powcoder.com>

Ratchet Accuracy

Add WeChat powcoder

The goal of the project is to achieve good model performance (which ever metric you use to measure this). Every step contributes towards this goal.

Set some simple benchmarks early on. Treat changes that you make as experiments that potentially increase accuracy.

Performance is a ratchet that can only move in one direction (better, not worse).

Adapt As Needed

Modify the steps as you need on a project, especially as you become more experienced with using the Notebook.

The final submitted Notebook does not need to preserve the suggested structure if you think something else is more appropriate.

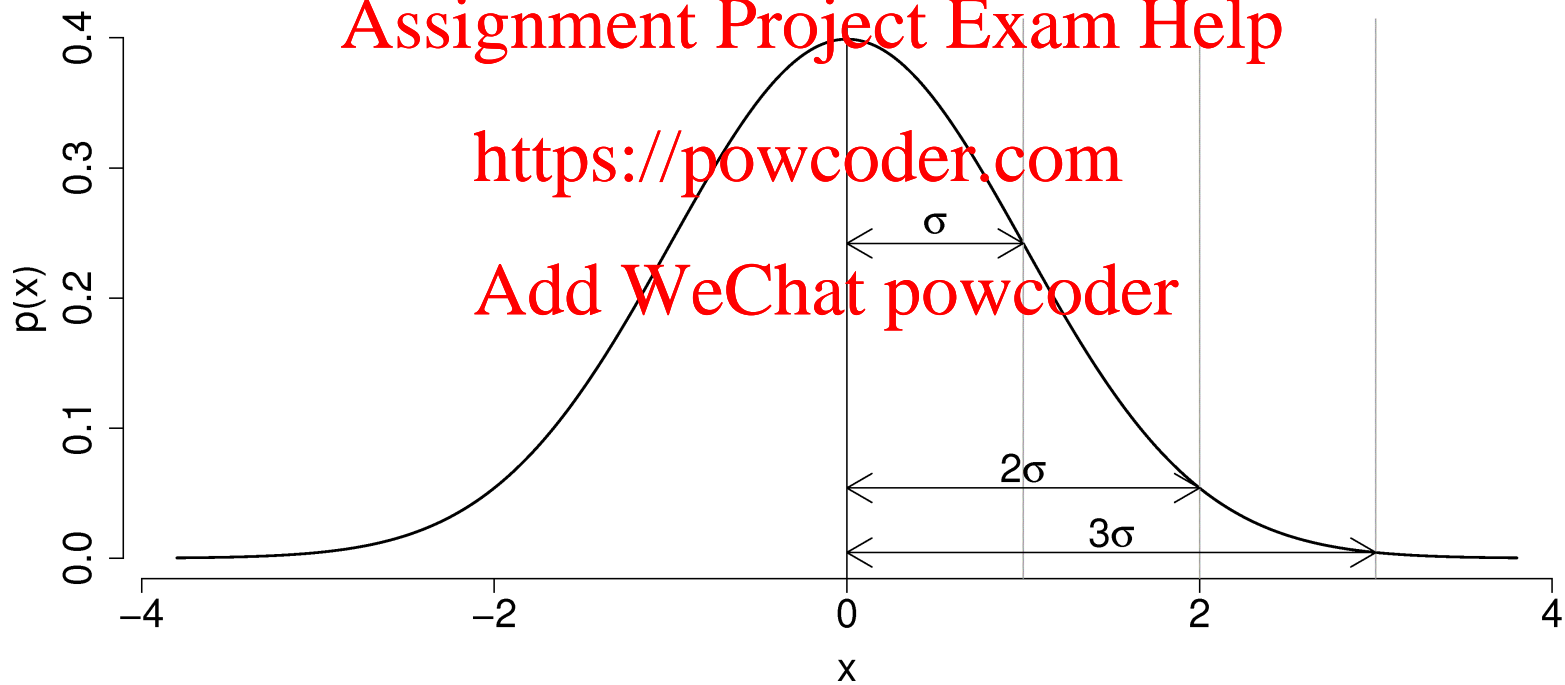
Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

A NOTE ON GRADES

Normal distribution when $\mu=0$ and $\sigma=1$



— Submission **Friday 26th February 2021, 10 am**

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Kamil Tylinski

Teaching Assistant

kamil.tylinski.16@ucl.ac.uk



Jiangbo Shangguan

Teaching Assistant

j.shangguan.17@ucl.ac.uk



Bartos Kultys

Teaching Assistant

bartosz.kultys.18@ucl.ac.uk



Editha Nemsic

Teaching Assistant

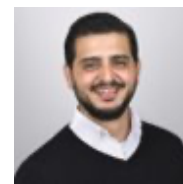
editha.nemsic.19@ucl.ac.uk



Dr Viviana Culmone

Teaching Assistant

v.culmone@ucl.ac.uk



Walter Hernandez

Teaching Assistant

walter.hernandez.18@ucl.ac.uk

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder