Capstone project idea 1

Can Wine quality predict based on physicochemical?

The data can be obtained at

<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>

Two datasets are included, related to red (1599 instances) and white vinho verde wine(4898 instances) samples from the north of Portugal.

The question that I try to answer are:

1. Are white and red wine quality predicted by the same physicochemical?
2. What the biggest factor that affect white or red wine quality?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 4898 | **Area:** | Business |
| **Attribute Characteristics:** | Real | **Number of Attributes:** | 12 | **Date Donated** | 2009-10-07 |
| **Associated Tasks:** | Classification, Regression | **Missing Values?** | N/A | **Number of Web Hits:** | 994017 |

Capstone project idea 2

Crime and safety are concerns for anyone wanting to travel, visit, or live in the city. Because of this, it is important to have an idea of what to expect regarding police reports that have occurred in the past.

This is a dataset containing records from the crime incident report from Boston, which include a reduced set of fields focused on capturing the type of incident as well as when and where it occurred.

This dataset has 2,60,760 rows and 17 columns. That includes:

INCIDENT\_NUMBER, OFFENSE\_CODE, OFFENSE\_CODE\_GROUP, OFFENSE\_DESCRIPTION, DISTRICT, REPORTING\_AREA, SHOOTING, OCCURRED\_ON\_DATE, YEAR, MONTH, DAY\_OF\_WEEK, HOUR, UCR\_PART, STREET, LATITUDE, LONGITUDE, LOCATION:

The questions that I try to answer are:

* How has crime changed over the years?
* Is it possible to predict where or when a crime will be committed?
* Which areas of the city have evolved over this time span?
* What time (maybe at night) most likely to have crime?
* In which area most crimes are committed?

<https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system>

|  |  |  |
| --- | --- | --- |
|  |  |  |

Capstone project idea 3

This data includes women tennis match data from 2000 to 2016, the question that I try to answer are:

* Who won the most title in each grand slam?
* Is it possible to predict who will win the women’s tennis match?
* Will age and ranking affect who win the game and so on?

The data are here:

<https://www.kaggle.com/gmadevs/wta-matches>