**INFM 600 – ASSIGNMENT 1 – INFORMATION SEEKING**

**Dataset 1** - **Bitcoin Historical Data**

**Source**

Kaggle. This dataset represents bitcoin data at 1-min intervals from select exchanges through the period of Jan 2012 to July 2018. Retrieved from <https://www.kaggle.com/mczielinski/bitcoin-historical-data> and API - Kaggle datasets download -d mczielinski/bitcoin-historical-data. Accessed on September 9, 2018

**Terms of Use**

Kaggle allows the users to use the resources through the before mentioned terms and restrictions.

“You will only use the Services for your own internal, personal, non-commercial use, and not on behalf of or for the benefit of any third party, and only in a manner that complies with all laws that apply to you. If your use of the Services is prohibited by applicable laws, then you aren’t authorized to use the Services. We are not responsible if you use the Services in a way that breaks the law. If you’re agreeing to these Terms on behalf of an organization or entity, you represent and warrant that you are authorized to agree to these Terms on that organization or entity’s behalf and bind them to these Terms, but you are still required to sign up each individual user from your organization with a Kaggle User ID.”

Link - <https://www.kaggle.com/terms>

**Description of Data**

The dataset presented describes about complete statistics and variations along with factors affecting the rise and fall of Bitcoin value in the international market and stock exchange. The data describes factors like opening value, closing value, high value, low value, weighted price, volume in currency and more about the bitcoin at an instant time. The data has an interval of 1 minute which allows to determine 24x60 changes of the bitcoin value throughout one day cycle. The dataset holds information of bitcoin value fluctuations starting from January 2012 to July 2018.

**About this file**

Timestamps are in Unix time. The data set with bitstamp in USD value has 3.41 million observations spread across 8 attributes which gives a huge range of data and opportunity to explore and understand patters of bitcoin variations throughout the period. Timestamps without any trades or activity have their data fields forward filled from the last valid time period. Below are the 8 attributes which would help determine inferences and analysis through the dataset: -

* Timestamp (in Unix time)
* Open- Bitcoin price in Currency units at time period open
* High- Highest Bitcoin price in Currency units during time period
* Low- Lowest Bitcoin price in Currency units during time period
* Close- Bitcoin price in Currency units at time period close
* Volume (BTC) - Volume of BTC transacted in time period
* Volume (Currency)- Volume of Currency transacted in time period
* Volume-weighted average price (VWAP)

**Potential data users and decision-makers for this data**

Investment Bankers, Financial institutions, Inside traders, Asset Management Firms, Consulting Management Firms, Multinational Banks, Stock Exchange Market and all the small and large scale business and firms who are continuously investing in the market or expecting returns from their investments as well as professionals who act as brokers for their clients to invest, all aforementioned are the potential data users as well as decision-makers for this data set.

There can be Data Analysts who can be appointed to serve as the decision-makers upon data analysis over these data and determine correct time to invest or withdraw investments for the investors or venture capitalists.

**Three Potential Questions this Data might help answer**

1. Observing data trends in the past, what can be the correct time to invest or withdraw investments?
2. Can Bitcoin with its trending patterns be utilized as a currency in regular civil use in the near future?
3. Determining the economic scenario of a particular currency, how did it affect the bitcoin patterns before and after the currency value fluctuations?

**Dataset 2** - **Abalone Data Set**

**Source –** University of California, Irvine – Machine Learning Repository. This dataset represents all physical and inherent attributes of abalones which can be utilized to understand and predict the age of abalones. The data was donated to the repository on 1995-12-0. Retrieved from <http://mlr.cs.umass.edu/ml/datasets/Abalone> and API - http://mlr.cs.umass.edu/ml/machine-learning-databases/abalone/abalone.data. Accessed on September 10, 2018

**Terms of Use**

UCI Machine Learning Repository – Center for Machine Learning and Intelligent Systems allows the users to use the resources through the before mentioned citation terms.

“If you publish material based on databases obtained from this repository, then, in your acknowledgements, please note the assistance you received by using this repository. This will help others to obtain the same data sets and replicate your experiments.”

Also, pseudo-APA reference format for referring to this repository: Asuncion, A. & Newman, D.J. (2007). UCI Machine Learning Repository [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science.

Link - <http://mlr.cs.umass.edu/ml/citation_policy.html>

**Description of Data**

The dataset presented describes about extensive list of measurements that can be obtained through an abalone starting from sex, length, diameter, eight, whole weight, shell weights and more. All of these measurements allow us to perform statistics and calculations and the results can be utilized to understand the growth patterns of an abalone and determine its age as well as life remaining for survival. Additional information like the location and weather patterns as well as food availability can be pictured into to have an in-depth analysis and improvise the precision of prediction.

**About this file**

The dataset file contains data with 1048576 observations with 8 attributes which allows us to perform an extensive analysis over utilizing a bunch of large data and the prediction keeps on improving as the sample size for the test increases. Below are the 8 attributes which would help determine inferences and analysis through the dataset: -

* Sex / nominal / -- / M, F, and I (infant)
* Length / continuous / mm / Longest shell measurement
* Diameter / continuous / mm / perpendicular to length
* Height / continuous / mm / with meat in shell
* Whole weight / continuous / grams / whole abalone
* Shucked weight / continuous / grams / weight of meat
* Viscera weight / continuous / grams / gut weight (after bleeding)
* Shell weight / continuous / grams / after being dried

**Potential data users and decision-makers for this data**

Scientists working on understanding the age, factors affecting health and survival for this particular species, research scientist utilizing the facts obtained for genetic purposes, zoologists intending to understand the life patterns and requiring information to substantiate the growth of abalones and preserving their extinction at a particular geographic environment can be a set of potential users and decision makers for this data.

**Three Potential Questions this Data might help answer**

1. How can we provide required environment and living conditions for abalones in areas where the analysis shows early ageing of abalones or findings which deviate from normal?
2. What can be the living patterns of abalones in a geographic region and conditions affecting their attributes?
3. Most world abalone populations have been overfished, some to virtual extinction. This dataset can be used to analyze patterns and conditions of healthy living of abalones and such locations can be encultured to make a suitable habitat for the abalones and create artificial conditions at other important locations.

**Dataset 3** - **Intra-Clutch Differences in egg characteristics**

**Source –** Dryad Digital Repository. This dataset represents Intra-clutch differences in egg characteristics which mitigate the consequences of age-related hierarchies in a wild passerine. Retrieved from <http://datadryad.org:8080/resource/doi:10.5061/dryad.8qj87> and API- https://doi.org/10.5061/dryad.8qj87. Published on April 15, 2013 and Accessed on September 11, 2018.

**Terms of Use**

Dryad Digital Repository allows the users to use the resources through the before mentioned terms and restrictions.

“Repository Users are allowed and encouraged to reuse Content from the Repository in any manner except as described herein under "Prohibited Uses Generally". To the extent possible under law, Submitters have waived all copyright and related or neighboring rights to this data. Users, Purchasers and Submitters may not use the Repository (i) in any unlawful manner; or (ii) in any manner that would impair Dryad's services to others or compromise the security or functionality of the Repository or Dryad's services generally.”

Link - <http://datadryad.org:8080/pages/policies#privacy>

**Description of Data**

The dataset presents Intra-clutch differences in egg characteristics which mitigate the consequences of age-related hierarchies in a wild passerine helping to understand the characteristics of eggs which are just laid as well as the eggs which are having different weights, exposure, characteristics which would be helpful to determine the time of eggs to hatch, differences and observations in the birth as well as fitness variation of species. Chicks that hatched earlier are heavier or lighter and had higher or lower probability of fledgling, and chicks that hatched from experimental eggs had patterns of growth and survival consistent with this. Egg mass is also explained a small part of this variation. These results are consistent with the idea that intrinsic differences between eggs across the laying sequence serve to mitigate the effects of age-related hierarchies.

**About this file**

The dataset file contains data with 2576 observations with 14 attributes which allows us to critically observe and understand the patterns and features of eggs and produce analysis report over the same. Below are the 14 attributes which would help determine inferences and analysis through the dataset :-

* egg\_weight
* nest\_orig
* hatched
* year
* XF
* flay
* rankfeB
* treatment
* nest\_rear
* hday
* min\_htime
* max\_htime
* incubation
* clutchsize

**Potential data users and decision-makers for this data**

Research organizations, individuals, poultry farmers, backyard poultry researchers, food chain researchers, industry R&D teams which are based on food products related to the animal produce, wildlife preservers and other similar profiles and organizations are appropriate and potential data users and decision makers utilizing this dataset.

**Three Potential Questions this Data might help answer**

1. Determination of healthy and marred eggs which might take up more resources and still not give birth to a healthy chicks?
2. How to proliferate life and produce of the species based on the existing observant conditions and attributes of the eggs and their characteristics?
3. What can be the best suitable environment for the reduced incubation period and healthy produce and births of the species?

Word Count - 1580