

RISHI JAIN: 116183031

## INFM 600 – ASSIGNMENT 1 – INFORMATION SEEKING

### 1. KAGGLE - Bitcoin Historical Data

Bitcoin data at 1-min intervals from select exchanges, Jan 2012 to July 2018

**Weblink** - <https://www.kaggle.com/mczielinski/bitcoin-historical-data>

**API** - Kaggle datasets download -d mczielinski/bitcoin-historical-data

#### About Data

Bitcoin is the longest running and most well known cryptocurrency, first released as open source in 2009 by the anonymous Satoshi Nakamoto. Bitcoin serves as a decentralized medium of digital exchange, with transactions verified and recorded in a public distributed ledger (the blockchain) without the need for a trusted record keeping authority or central intermediary. Transaction blocks contain a SHA-256 cryptographic hash of previous transaction blocks, and are thus "chained" together, serving as an immutable record of all transactions that have ever occurred. As with any currency/commodity on the market, bitcoin trading and financial instruments soon followed public adoption of bitcoin and continue to grow. Included here is historical bitcoin market data at 1-min intervals for select bitcoin exchanges where trading takes place. Happy (data) mining!

#### About this file

CSV files for select bitcoin exchanges for the time period of Jan 2012 to July 2018, with minute to minute updates of OHLC (Open, High, Low, Close), Volume in BTC and indicated currency, and weighted bitcoin price. Timestamps are in Unix time. Timestamps without any trades or activity have their data fields forward filled from the last valid time period. If a timestamp is missing, or if there are jumps, this may be because the exchange (or its API) was down, the exchange (or its API) did not exist, or some other unforeseen technical error in data reporting or gathering. All effort has been made to deduplicate entries and verify the contents are correct and complete to the best of my ability, but obviously trust at your own risk.

- 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)

## Columns

- Timestamp
- Open
- High
- Low
- Close
- Volume\_(BTC)
- Volume\_(Currency)
- Weighted\_Price

## PRIVACY POLICY

When you clicked “accept” or “agree” in connection with registering your account, we made available to you this Privacy Policy and the information handling practices described in here. This Privacy Policy only covers Kaggle's practices regarding personal information, and not information more generally. We may use aggregated, anonymized data that we derived from your personal information before you deleted it, but not in a manner that incorporates any of your personal information or would identify you personally.

We'll ask for your consent before using your information for a purpose that isn't covered in this Privacy Policy. You may, of course, decline to share certain personal information with Kaggle, in which case Kaggle may not be able to provide to you some of the features and functionality found on the Services.

**LINK:** <https://www.kaggle.com/privacy>

## 3 QUESTIONS DATA MIGHT BE ABLE TO ANSWER

- Trends of Cryptocurrency
- Rising Patterns and Peaks of Bitcoin Value
- Predications of future rates of Bitcoin.

## 2. Abalone Data Set

**Weblink:** <http://mlr.cs.umass.edu/ml/datasets/Abalone>

**API** - <http://mlr.cs.umass.edu/ml/machine-learning-databases/abalone/abalone.data>

## ABOUT DATA

Predicting the age of abalone from physical measurements. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope -- a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age. Further information, such as weather patterns and location (hence food availability) may be required to solve the problem.

From the original data examples with missing values were removed (the majority having the predicted value missing), and the ranges of the continuous values have been scaled for use with an ANN (by dividing by 200).

## **COLUMNS**

Given is the attribute name, attribute type, the measurement unit and a brief description. The number of rings is the value to predict: either as a continuous value or as a classification problem.

Name / Data Type / Measurement Unit / Description

- 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
- Rings / integer / -- / +1.5 gives the age in years

## **DATA OWNERS/CREDITS**

Data comes from an original (non-machine-learning) study:

Warwick J Nash, Tracy L Sellers, Simon R Talbot, Andrew J Cawthorn and Wes B Ford (1994)  
"The Population Biology of Abalone (*Haliotis* species) in Tasmania. I. Blacklip Abalone (*H. rubra*) from the North Coast and Islands of Bass Strait",  
Sea Fisheries Division, Technical Report No. 48 (ISSN 1034-3288)

Original Owners of Database:

Marine Resources Division  
Marine Research Laboratories - Taroona  
Department of Primary Industry and Fisheries, Tasmania  
GPO Box 619F, Hobart, Tasmania 7001, Australia  
(contact: Warwick Nash +61 02 277277, [wnash '@' dpi.tas.gov.au](mailto:wnash@dpi.tas.gov.au))

Donor of Database:

Sam Waugh ([Sam.Waugh '@' cs.utas.edu.au](mailto:Sam.Waugh@cs.utas.edu.au))  
Department of Computer Science, University of Tasmania  
GPO Box 252C, Hobart, Tasmania 7001, Australia

## PRIVACY POLICY

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. We suggest the following pseudo-APA reference format for referring to this repository:

LINK: [http://mlr.cs.umass.edu/ml/citation\\_policy.html](http://mlr.cs.umass.edu/ml/citation_policy.html)

## 3 QUESTIONS DATA MIGHT BE ABLE TO ANSWER

- Age of Abalone
- Traits of Abalones
- Predications of patterns of Abalones.

## 3. Intra-Clutch Differences in egg characteristics

Weblink: <http://datadryad.org:8080/resource/doi:10.5061/dryad.8qj87>

API - <https://doi.org/10.5061/dryad.8qj87>

## ABOUT DATA

Data from: Intra-clutch differences in egg characteristics mitigate the consequences of age-related hierarchies in a wild passerine.

The relative age of an individual's siblings is a major cause of fitness variation in many species. In Blue tits (*Cyanistes caeruleus*) we show that age hierarchies are predominantly caused by incubation pre-clutch completion, such that last laid eggs hatch later than early laid eggs. However, after statistically controlling for incubation behavior late laid eggs are shown to hatch more quickly than early laid eggs reducing the amount of asynchrony. By experimentally switching early and late laid eggs between nests on the day they were laid we controlled for the effect of differential incubation and found that the faster hatching times of late laid eggs remains.

Chicks that hatched earlier were heavier and had higher probability of fledgling, and chicks that hatched from experimental eggs had patterns of growth and survival consistent with this. Egg mass explained a small part of this variation, but the remainder must be due to egg composition. 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. We also show that between-clutch variation in prenatal developmental rate exists and that it is mainly environmental in origin rather than genetic.

## **COLUMNS**

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

## **DATA OWNERS/CREDITS**

When using this data, please cite the original publication:

Hadfield JD, Heap EA, Bayer F, Mittell EA, Crouch NMA (2013) Intra-clutch differences in egg characteristics mitigate the consequences of age-related hierarchies in a wild passerine. *Evolution* 67(9): 2688-2700. <https://doi.org/10.1111/evo.12143>

Additionally, please cite the Dryad data package:

Hadfield JD, Heap EA, Bayer F, Mittell EA, Crouch NMA (2013) Data from: Intra-clutch differences in egg characteristics mitigate the consequences of age-related hierarchies in a wild passerine. Dryad Digital Repository. <https://doi.org/10.5061/dryad.8qj87>

**LINK:** <http://datadryad.org:8080/resource/doi:10.5061/dryad.8qj87/1>

## **PRIVACY POLICY**

By submitting content to or using content from the Dryad Digital Repository, submitters and users enter into a binding agreement to accept the Terms of Service described herein. If you do not agree to these Terms of Service, please refrain from using content from or submitting content to the Repository.

Guiding these Terms of Service are Dryad's aims to:

- enhance the scholarly record by making data freely available and reusable for research and education;
- enable repository content to be widely indexed and discoverable;
- assign and provide Digital Object Identifiers (DOIs) to repository content;
- track the usage of content, and promote data reuse as a metric of scholarly credit;
- manage data publication embargoes as stipulated by publisher policies;
- make data both accessible today and preserved for accessibility in the future;
- respect and protect the privacy of its users; and
- generally, meet the needs of its stakeholders for a trustworthy repository.

**LINK:** <http://datadryad.org:8080/pages/policies>

### **3 QUESTIONS DATA MIGHT BE ABLE TO ANSWER**

- Intra-clutch Variations in Eggs.
- Statistics behind egg hatching.
- Predictions of egg hatching on simulated environments.