

## COMS30038 — SECURITY BEHAVIOURS

### LAB 4: DARKNET MARKET SURVIVAL

---

In this session, you will be working with data summarising the histories of darknet marketplaces involved in the English-language online drugs trade in the 2010s, using a dataset that chronicles their service duration and known features. The aim is that working with the dataset will give you some deeper insight into the nature of this type of cybercriminal economy, and highlight some security issues exacerbated by the incentive structure in these marketplaces.

#### 1 FACT-FINDING

For this exercise you can work alone, or in a pair or small group. You will need to begin by accessing the `dnm_histories.csv` file in your preferred data exploration tool (e.g., R or Python). This is a public dataset created by an independent researcher, recording the start and end of operations for 88 darknet marketplaces, and some features of those marketplaces<sup>1</sup>. You may want to start by understanding the structure of the dataset. Once you have, try to answer the following questions:

1. What was the most common cause of a marketplace closure?
2. What percentage of “doxed” market owners were also arrested?
3. Which market survived the longest?
4. How many markets survived longer than Silk Road 1?
5. Generously defining ‘success’ as a market surviving for an entire year, how many successful darknet markets were there?
6. Which group of markets suffered the greater rate of hacks, the ‘successful’ or the ‘unsuccessful’ ones, per the definition above?

---

<sup>1</sup>See <https://gwern.net/dnm-survival> for more information on the dataset’s creation, and for an explanation of each of the columns.

7. Were markets offering multisig more likely to be successful than those that did not?
8. Permitting which of the two restricted goods categories (guns / proceeds of fraud) has the strongest association with market success?

## 2 SURVIVAL ANALYSIS

While it has other uses, this dataset was originally created with the aim of measuring the relationship of market features to longevity. Continuing in your group or alone, try to (a) identify which individual factors in the dataset have the strongest association with a market's lifespan and (b) see how well you can fit a predictive model using a subset of the available features, and measure its accuracy. This is an open task which your TA can help you explore but without a specific solution. Try to relate what you find back to what this implies about the incentives for and against different approaches to security in cybercriminal markets.

To address individual factors, research an appropriate approach for measuring the relationship between the factor type (most but not all of the factors are binary) and longevity of a market (which can be measured as a count of days). Note that while similar, this is not the same as the binary 'success' decision we applied in several fact-finding questions above.

When it comes to the predictive model, you can dive as deep as you like into this topic. A simple linear model would be a good first effort, but you may run into some problems when incorporating certain features. If you are interested, you can research and try to apply specific survival-analysis models to this dataset using open statistical libraries.

---

### 3 WRITING PRACTICE

This section isn't part of the lab itself, but an opportunity for you to practice writing essay-style answers to (week-topical) questions. You can show your answer to your TA next week, or share in your group's Teams channel to get feedback on how you're answering questions.

1. Give an example of a situation where the concept of a 'lemon market' applies, and argue to justify the case that a lack information access is driving down prices in your example. [≈250 words]
2. Discuss the importance of trust in cybercriminal markets. Is trust more or less important than in legitimate markets? [≈400 words]