Ruqayyah Muse

Dr. Jean-Paul Addie

Honors 3260

10 October 2023

# The Intersection of Data Science and the Study of Cities at Night

In 2019, New York's nightlife accounted for $29 billion of economic activity and provided around 250,000 jobs (Acuto). Similarly, in London, an estimated one-third of the population's workers worked an evening or night shift as of 2018 (Acuto; "London at Night: An Evidence Base for a 24-Hour City"). Although those numbers likely shrunk due to the COVID-19 pandemic, this data shows how large and impactful the night-time economy of cities is, and how it can have a vast effect on individuals and society. This has led to calls for the establishment of an interdisciplinary night studies field that encourages collaboration between different disciplines to solve the multi-faceted issues that arise with regard to the city at night. One such discipline that can be useful is data science, a field that combines statistics, programming, analytics, and artificial intelligence to find relationships and insights hidden in data ("What Is Data Science?"). This can make it a useful tool for understanding the social, economic, and health issues surrounding the city at night. In this paper, I will explore how data science can support the research of cities at night, consider relevant applications and questions that can be answered, and share the limitations of data science in this kind of research.

## WHAT IS DATA SCIENCE AND HOW IS IT APPLIED

### *THE DATA SCIENCE LIFECYCLE*

In order to understand how data science can be applied to solve problems of the night, it's important to first understand how the data science lifecycle works. One of the most common lifecycle frameworks is CRISP-DM, which gives a high-level overview of what data science essentially is (Nantasenamat). With the CRIPS-DM framework, there are 6 steps: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Since deployment, which refers to building a report, or a mobile/web application, does not occur in every case, I will be focusing on just the first five steps of the framework, starting with business understanding.

In the business understanding stage, the goal is to define the scope of the problem that needs to be solved. Some examples of business problems that data science has been used for include predicting the likelihood of employee attrition or understanding the sentiment behind customer reviews of products. It can also be used to answer social or economic questions, such as understanding the correlation between race and household income. In any case, understanding the scope of the problem to be solved is important and would determine the type of data and modeling that would be used in future stages.

Once the problem has been defined, the next step is data understanding, which includes gathering the data. Data can be of all types and forms, such as time-series data, graph data, textual data, image data, etc. so long as it is relevant to the problem at hand. At times, there might not be existing data to use, so data scientists would set up surveys or begin a data collection process to gather the data that they need. Also included in data understanding is performing exploratory data analysis. This would include visualizing the distribution of the data, and finding out whether there are missing values, or if there are any immediately noticeable subsets that need to be handled.

After the business problem and the data are understood, the data is ready to be processed and modeled. During the processing step, the data is cleaned and formatted by handling missing or null values, removing unnecessary attributes, normalizing the values within certain ranges, and any other processing steps that are needed to prepare it for the modeling stage. The modeling stage includes choosing a model, which is essentially an algorithm, that can solve a business problem. Once an appropriate model is selected, the model is trained on a portion of the data so that it can learn how to make classifications and predictions based on the information it is given. The model is then tested against the remaining data in order to fine-tune the accuracy and ensure that the model produces the desired results.

The final step is the evaluation or interpretation of the results. Once we have a working model, we need to be able to understand and draw meaningful conclusions based on what the model is telling us. For instance, if we performed sentiment analysis on customer reviews, the model would likely return a numeric value corresponding to how positive or negative a review was. Being able to use those numeric values to classify the reviews and find the common factors that caused a certain sentiment is an example of how interpreting the results can yield a meaningful conclusion. This concludes the general stages that are necessary for the data science lifecycle.

### *Applications of Data Science*

Due to the wide range of what data science as a discipline covers, something as simple as creating a graph can fall under the umbrella of data science as much as building an artificial intelligence model that predicts or creates new information does. This means that there are many applications for data science, such as the recently famous language learning model ChatGPT. Other examples that are not directly noticeable to the public include resume-review programs that filter out certain candidates based on a given job description. Similarly, facial recognition, email spam filtering,

and even criminal justice systems used to determine the likelihood of recidivism are all ways that data science has been applied. Given the versatility of the discipline, it would be extremely useful to use data science for studying the night, and for developing policies that cities can use to better govern the night.

### *The Intersection of Data Science with Public Policy*

One interdisciplinary way that data science has already been used for city planning is through public policy work. The Data Science and Public Policy Lab at Carnegie Mellon University has been partnering data scientists with local governments in order to address policy issues with regard to housing, transportation, education, and other sectors. One such project shows how data science was used to build a Transit Equity Visualizer, "a reusable software tool for evaluating whether changes in [public transportation] service adversely impact minorities and other disadvantaged individuals" in the city of Seattle in Washington state (Naveed et al.). With this example, we can see how data science is already being used to solve day-time city policy issues, which shows that it could be useful in solving night-time policy issues as well. For instance, this model could be extrapolated to evaluate the efficiency of night-time public transportation to determine who really needs such services and whether they are receiving them.

## How Can Data Science Contribute to Understanding the City at Night

### *Crime and the Fear of the Night*

Societal distrust of the night is a concept that can be traced back well into history. This distrust stemmed mostly from the lack of lighting, thus fueling the perception that the night is used as a cover for debauchery and crime. In one description from the 18$^{th}$ to 19$^{th}$ century prior to the Industrial Revolution, the night was said to be "a forbidding place plagued by pestilential vapors, diabolical spirits, natural calamity and human depravity" (Ekirch 56). With the introduction of gas

and electric lighting in the 19th century, the perception of the night slowly began to change, with lighted streets being seen as a sign of prosperity, and prestige (Kenny). However, this advancement in technology still did not eradicate the negative perceptions associated with the night, and that has continued to the modern day. Even with our more reliable and consistent city lights, the night is still perceived by certain demographics, especially women and minorities, as a dangerous place to be alone (Dugan). This could be due in part to America's perception that crime is still increasing, as well as due to news stories highlighting kidnappings, murders, and other violent crimes that have taken place at night.

One way that data science can be useful in this area is by analyzing the data we have collected on crime in cities at night. There are several questions that data analysis and modeling can answer, such as: what kind of crimes are more likely to occur at night; do crimes tend to occur in urban parts of the city near night-time establishments; are there certain times of night that are more dangerous than others; and what common factors are found when analyzing the types of crimes reported at night. Many of these questions can be answered by using data from police incident reports, or self-reported crime surveys. There have already been some studies trying to analyze this data, such as a recent 2023 study that sought to find factors that contribute to crime in the night-time economy (Wüllenweber and Burrell). By using data science with these questions, city governments would be better informed on which policies might increase people's positive perceptions of the night, thus making the night feel safer.

*BUILDING A STRONG NIGHT-TIME ECONOMY*

The night-time economy encompasses all the industries that work predominantly in the night, such as the entertainment, hospitality, maintenance, and logistics industries. By studying cities such as New York, London, Amsterdam, and others, we can gather relevant data to see what makes the night-time economy in those cities so successful to replicate that success in other cities, such as

here in Atlanta. We can also use that data to solve problems faced by people who work in the night-time economy, such as determining what kind of transportation at what times is most useful for workers commuting at night. Similarly, we could gather data on the effects of lighting and advertisements at night and see if those have any effect on people's desire to be in the city at later times.

In order to answer these economic questions, we can utilize economic and labor statistics gathered by cities. For instance, there was a 2018 "London at Night: An Evidence Base for a 24-Hour City" report made by the city of London which gathered data on night-time industries, night-time workers, transportation, and people who participated in the night-time economy. Another method for gathering data could be through random population surveys where data scientists can directly ask people what it is that makes them want to be in the city at night. By using these types of sources, we can better understand what it is that makes a successful and enjoyable night-time economy for all parties involved.

**LIMITATIONS OF DATA SCIENCE**

Although data science can be a versatile discipline with various applications and the ability to answer questions pertaining to the city at night, it does not come without its limitations. One of the biggest limitations is finding plentiful, useful data. Since the study of the night is a relatively new field, data scientists may lack detailed collections of data making it more difficult to answer some of the questions posed. For instance, many cities keep track of work and employment statistics, but these statistics might not always take into consideration the time of day when this work occurs, making it harder to study night worker statistics separate from day workers. Nonetheless, there are methods for data scientists to synthesize or extract some useful information, but it does make it more difficult to confidently conclude results when there is not enough data to support the findings.

Another limitation of data science comes from the evaluation of results. Despite the perceived objectivity of using a program to predict information, the results may not be necessarily accurate. Data science is as much a human-driven science as it is a machine-driven one. A data scientist chooses the data that is fed to the model, which in turn influences the conclusion that the model makes. Using biased data is one way that can result in inaccurate models, as has happened in the case of facial recognition systems that have been found to be less accurate on darker-skinned women (Guttag and Suresh). A less accurate or biased model can also occur due to there being subsets within the data that the model can not interpret, which would result in skewed predictions (Guttag and Suresh). Thus, the proposal of a certain conclusion based on a data science experiment or model still needs to be analyzed and cross-referenced so that the right policies can be implemented for cities at night.

## CONCLUSION

In conclusion, data science can be a useful lens with which to approach the study of the city at night. Given its scientific empirical nature, it can provide a more objective way of making policy decisions while at the same time yielding measurable results for future analysis. Data Science is already being implemented in city planning, which also sets a precedent that supports its use for night-time city planning as well. As we start to use data science to support the field of night studies, it is also important to be aware of the limitations of data science, including how the lack of data and the potential for data to be biased or incomplete can lead to potentially biased predictions and results. Nonetheless, data science has the potential to be a driving force in the field of night studies.

Works Cited

Acuto, Michele. "We need a science of the night." *Nature*, vol. 576, no. 7787, 18 Dec. 2019, pp. 339–339, https://doi.org/10.1038/d41586-019-03836-2.

Dugan, Andrew. "In U.S., 37% Do Not Feel Safe Walking at Night near Home." *Gallup.Com*, Gallup, 15 Aug. 2022, www.news.gallup.com/poll/179558/not-feel-safe-walking-night-near-home.aspx.

Ekirch, A. Roger. *At Day's Close: A History of Nighttime*. Weidenfeld & Nicolson, 2005.

Kenny, N. (2017). City Glow: Streetlights, Emotions, and Nocturnal Life, 1880s–1910s. Journal of Urban History, 43(1), 91-114. https://doi.org/10.1177/0096144215576716.

"London at Night: An Evidence Base for a 24-Hour City." *London City Hall*, Greater London Authority, 8 Nov. 2018, www.london.gov.uk/sites/default/files/london-at-night-full-final.pdf.

Nantasenamat, Chanin. "The Data Science Process." *Medium*, Towards Data Science, 14 Jan. 2022, www.towardsdatascience.com/the-data-science-process-a19eb7ebc41b.

Naveed, Hareem, et al. "Measuring and Increasing Equity in Transportation Access in Seattle." *Data Science and Public Policy*, Carnegie Mellon University, 1 Feb. 2023, www.datasciencepublicpolicy.org/portfolio-items/madison-rapidride-network-ridership-and-equity-analysis/.

Suresh, Harini, and John V Guttag. "A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle." ArXiv.org, ArXiv, 1 Dec. 2021, https://arxiv.org/abs/1901.10002.

"What Is Data Science?" *IBM*, www.ibm.com/topics/data-science.

Wüllenweber, S., & Burrell, A. (2023). The crime and the place: Robbery in the night-time economy. *Journal of Investigative Psychology and Offender Profiling*, 1–17. https://doi.org/10.1002/jip.1616.