The Journal of Open Source Software

hillmaker: A Python package for occupancy analysis in discrete entity flow systems

Mark W. Isken [1] **and Jacob W. Norman**[2]

1 Oakland University, USA **2** UNC Health Rex, USA

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Marcel Stimberg

Reviewers:

- @rerickson-USGS
- @HLasse

Submitted: 16 November 2023
Published: unpublished

Summary

hillmaker is a Python package that computes time of day and day of week specific arrival, departure, and occupancy statistics from transaction data containing arrival and departure timestamps. Typical use is for capacity planning problems in places like hospital emergency departments, surgical recovery rooms or any system in which entities arrive, occupy capacity for some amount of time, and then depart. It gets its name from the hill-like nature of summary occupancy plots - see Figure 1.

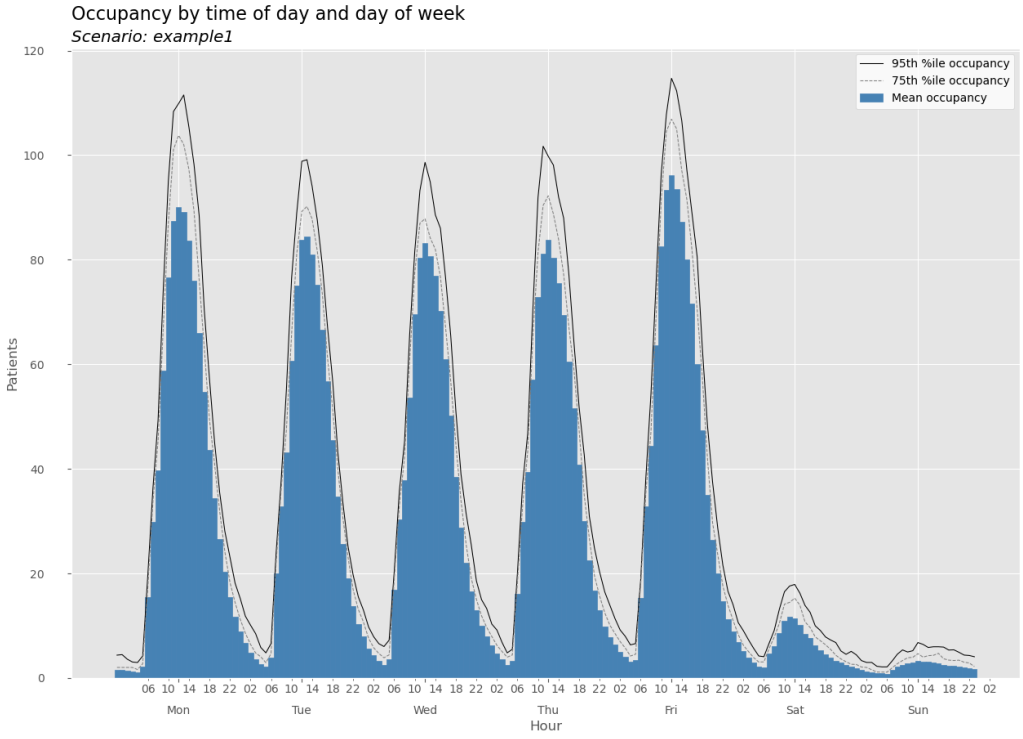

Figure 1: Weekly occupancy plot.

The hillmaker package can be used as a command line application as well as an importable library from Jupyter notebooks or Python scripts. There is an object-oriented API as well as a function based API. Under the hood, hillmaker relies primarily on numpy (Harris et al., 2020), pandas (team, 2020), matplotlib (Hunter, 2007), seaborn (Waskom, 2021) and Pydantic. Transaction data can be in CSV format or a pandas `DataFrame`. The primary outputs of hillmaker are:

- pandas `DataFrames` and CSV files with arrival, departure, and occupancy summaries by time of day, day of week and, optionally, an entity category.
- customizable weekly and daily plots of arrivals, departures, and occupancy by time of day and day of week.
- summary tables and plots for length of stay and implied operating hours.

The documentation includes several tutorials on using hillmaker for typical occupancy analysis problems in healthcare.

Statement of need

The original motivation for hillmaker was a series of capacity planning problems faced by a management engineering group at a large healthcare system. Such problems are characterized by patient flow and capacity use that exhibit significant and important time of day and day of week patterns. For example, a huge component of hospital costs are related to labor. Staffing and scheduling relies on effectively matching capacity to demand that varies significantly by time of day and day of week. Another motivating problem involved developing a surgical patient rerouting plan to temporarily accommodate patients while a post-surgical holding area renovation project was underway. Proper statistical analysis of patient arrival, departure, and occupancy patterns, including computation of percentiles, are a critical part of such analyses. Traditional statistics packages do not include this type of functionality and one-off SQL based approaches are tedious to create and generalize. Additionally, the implementation of percentile functions in SQL is dialect dependent.

Early versions of hillmaker were used in hundreds of projects in multiple healthcare engineering departments and consulting firms. It was written in Microsoft Access and released as an open source project in the early 2000s. You can still find it on SourceForge at https://sourceforge.net/projects/hillmaker/. An academic paper about this early version was published (Isken, 2002) after the first author left industry and joined Oakland University. Unfortunately, development on this version languished for a variety of technical and logistical reasons. So, hillmaker continued to get significant use but few improvements were made.

In addition to industry use, hillmaker is well-suited for healthcare operations research projects involved patient (Broyles et al., 2010; Helm & Van Oyen, 2014; Isken et al., 2011; R. Konrad et al., 2012), material (Isken & Littig, 2002), and even information flow (R. A. Konrad et al., 2008). Discrete event simulation is widely used in such research projects and hillmaker can aid in the statistical analysis needed for modeling entity arrival patterns and for analyzing and validating simulation output. While problems in the healthcare industry spurred the development of hillmaker, it has been used in other domains such as bike share systems, freight operations (Castrellon et al., 2023), customer contact centers, and even for analyzing usage patterns of a high performance computing cluster by engineers at a large automobile manufacturer. Any system for which you have data on start and stop times of events, or entry and exit times of entities, is amenable to using hillmaker for characterizing temporal patterns in arrivals, departures, and occupancy (or task starts, task completions and work in progress).

Given the rise of Python in the scientific computing world, rewriting hillmaker in Python made a lot of sense for its future as an open source project. The first Python version was released in 2016. A series of improvements followed over the next several years, culminating in the recent release of version 0.8.0 in November of 2023. Many of the recent improvements were motivated by the second author's use of hillmaker at UNC Health Rex, part of a larger healthcare system in North Carolina. They used hillmaker to better understand patient throughput and capacity management in their surgical services division as well as to determine the optimal bed size for a potential same day recovery unit.

By creating an easy to use Python version with extensive documentation, we are hoping that hillmaker can gain greater traction now that Python has become so widely used in the data analysis community. Our plan is to add a form-based GUI (much like the one in the

original MS Access version) in the next release which should make it even easier to use for non-programmers.

Acknowledgements

We acknowledge contributions from the many analysts and healthcare institutions for both their direct and indirect support of the ongoing evolution of hillmaker over the past three decades.

References

Broyles, J. R., Cochran, J. K., & Montgomery, D. C. (2010). A statistical markov chain approximation of transient hospital inpatient inventory. *European Journal of Operational Research*, *207*(3), 1645–1657. https://doi.org/10.1016/j.ejor.2010.06.021

Castrellon, J. P., Sanchez-Diaz, I., & Kalahasthi, L. K. (2023). Enabling factors and durations data analytics for dynamic freight parking limits. *Transportation Research Record*, *2677*(2), 219–234. https://doi.org/10.1177/03611981221115086

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Helm, J. E., & Van Oyen, M. P. (2014). Design and optimization methods for elective hospital admissions. *Operations Research*, *62*(6), 1265–1282. https://doi.org/10.2139/ssrn.2437936

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Isken, M. W. (2002). Modeling and analysis of occupancy data: A healthcare capacity planning application. *International Journal of Information Technology & Decision Making*, *1*(04), 707–729. https://doi.org/10.1142/s0219622002000439

Isken, M. W., & Littig, S. J. (2002). Simulation analysis of pneumatic tube systems. *Journal of Medical Systems*, *26*, 9–19. https://doi.org/10.1023/A:1013034719088

Isken, M. W., Ward, T. J., & Littig, S. J. (2011). An open source software project for obstetrical procedure scheduling and occupancy analysis. *Health Care Management Science*, *14*, 56–73. https://doi.org/10.1007/s10729-010-9141-8

Konrad, R. A., Lawley, M. A., Yih, Y., & Lambert, S. (2008). Characterizing inpatient hospital flow from information systems messaging. *IIE Annual Conference. Proceedings*, 1049.

Konrad, R., Kucukyazici, B., & Lawley, M. (2012). Using patient flow to examine hospital operations. In *Management engineering for effective healthcare delivery: Principles and applications* (pp. 402–417). IGI Global. https://doi.org/10.4018/978-1-60960-872-9.ch019

team, T. pandas development. (2020). *Pandas-dev/pandas: pandas* (latest). Zenodo. https://doi.org/10.5281/zenodo.3509134

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021