

Forecasting Jail Bed Utilization for Central Virginia Regional Jail

Bella Lu¹, Rachel Seo¹, Jonas Lee¹, Loreto Peter Alonzi III¹, Michael Smith¹, K. Preston White Jr.¹, Senior Member Matthew Vitale², and Superintendent Frank E. Dyer³

¹ *Criminal Justice Data Training Initiative*
UVA School of Data Science
Charlottesville, VA
{fpq5wq, ydp7xy, uhz7mq, lpa2a, mcs5f, kpw8h}@virginia.edu

² *Offender Aid and Restoration Office*
Jefferson Area Community Corrections
Charlottesville, VA
mvitale@oar-jacc.org

³ *Central Virginia Regional Jail*
Virginia Association of Regional Jails
Orange, VA

Abstract—The Central Virginia Regional Jail (CVRJ) comprises five jurisdictions, and must decide whether to add a sixth, Culpeper County. While the CVRJ currently maintains excess capacity, the inclusion might induce future overcrowding. In response to a direct request from the Superintendent, this study presents a ten-year forecast of jail bed utilization to support long-term capacity planning at the CVRJ and to assess the feasibility of Culpeper County joining CVRJ as a member jurisdiction. For nearly two decades, CVRJ has relied primarily on a model derived from financial data to guide decisions on capacity planning and facility expansion. Although this approach has guided previous planning efforts, current trends in jail utilization and inmate populations reveal that financial indicators alone may not fully reflect operational demand. To address the limitations, this study employs a Seasonal Autoregressive Integrated Moving Average with exogenous variables (SARIMAX) time-series framework that integrates crime and jail booking data to produce a more precise estimate of inmate length of stay and overall bed utilization. We find that under the scenario in which Culpeper joins CVRJ, the combined Average Daily Population is projected to reach 569 beds by 2031 and 584 beds by 2035, with daily exceedance probabilities of 17.6% and 27.2% respectively, against the facility’s maximum capacity of 660 beds. These findings indicate that CVRJ can accommodate Culpeper County within the next ten years, while underscoring the importance of continued data collection to refine long-term projections.

Keywords—criminal justice, Culpeper, CVRJ, jail population forecasting, SARIMAX, time series

I. INTRODUCTION

Law enforcement agencies increasingly rely on data analytics to optimize resource allocation and enhance public safety outcomes. In alignment with this broader trend, the Criminal Justice Data Training Initiative (CJDTI) at the University of Virginia School of Data Science has undertaken an investigation into the application of statistical modeling and visualization techniques for forecasting jail bed utilization.

This project serves two primary objectives: (1) to project future jail bed utilization at the Central Virginia Regional Jail (CVRJ) using historical population trends, and (2) to provide projections that inform operational planning and resource allocation decisions for Culpeper County administrators.

The resulting model addresses a demonstrable gap in data analysis capacity within CVRJ. Through partnership with the CJDTI, facility administrators gain access to advanced analytical capabilities that expand the range of evidence-based strategies previously unexplored. By translating raw data into an intuitive, visually accessible dashboard, this project bridges the divide between complex statistical outputs and actionable intelligence for institutional decision-makers.

We find that the combined Average Daily Population under a full Culpeper membership scenario remains below the 660-bed maximum capacity threshold throughout the forecast window. Our projections indicate that CVRJ will retain a comfortable operational buffer through 2031, with utilization gradually approaching peak capacity by 2035. The following sections detail the data sources, modeling pipeline, and findings that support this conclusion.

II. PROCEDURE

A. Data Acquisition

The datasets used in this analysis were obtained through a series of institutional partnerships. The data primarily originated from three distinct sources: Superintendent Dyer, Lieutenant Taylor McDaniel, and Criminal Justice Planner Matthew Vitale.

First, the Average Daily Population (ADP) data for Culpeper County inmates provided by Dyer spanned from 2014-2025. This dataset offered a jurisdiction-specific longitudinal record of inmate population trends, making it particularly valuable for projecting Culpeper’s future bed needs.

Second, a de-identified booking dataset for the CVRJ was provided by Lieutenant McDaniel, which contained individual-level records including booking dates, release dates, and county codes.

Third, the Virginia Beyond 20/20 Census data platform, provided by Vitale, showed census estimates from 2012-2025. This source supplied the regional demographic context necessary to situate jail utilization trends within broader population growth patterns across the CVRJ service area.

B. Description of Data

The de-identified CVRJ booking dataset served as the primary source for constructing the daily jail census. Each record captured the start and end of an individual’s incarceration period, along with the county code identifying the jurisdiction responsible for the inmate. As a result of the de-identified nature, no personally identifiable data was retained, ensuring that privacy standards were maintained.

The Culpeper ADP data provided annual aggregate population figures about Culpeper inmates held within their own jail and across all jail facilities, including Rappahannock-Shenandoah-Warren (RSW), another regional jail. This distinction is important as Culpeper currently operates under a Memorandum of Understanding (MOU) with CVRJ — a contractual arrangement in which Culpeper pays CVRJ on a fee-for-service basis to house its inmates without holding permanent membership in the facility — rather than being a permanent member jurisdiction. The VA Beyond 20/20 Census data complemented both datasets by providing annual county population estimates, which offered insight into how regional demographic growth may drive future increases in jail utilization. Together, these three sources provided operational detail and broader demographic context needed to build a robust forecasting model. Table I includes all the significant variables used in the model.

TABLE I
DATA DICTIONARY OF THE MOST PREVALENT VARIABLES

Source	Field Name	Description (Data Type)
De-identified	Book Date	Date of booking into CVRJ (Date)
De-identified	Release Date	Date of release from CVRJ (Date)
De-identified	County Code	Code of county of arrest (Int)
VA Beyond	County Population	Number of people in county from 2012–2025 (Int)
Culpeper Jail	Annual ADP	Average Daily Population for Culpeper County from 2014–2025 (Int)

C. Data Processing and Analysis

During evaluation of the de-identified booking dataset, the original Length of Stay (LOS) variable was found to exclude the year from its computation, resulting in incorrect stay durations for inmates whose incarceration spanned multiple calendar years. For example, a booking from October 1st, 2022 to October 2nd, 2023 would be recorded as one day rather than 366 days. This error was identified and the LOS values were recalculated before any further analysis was conducted, ensuring that subsequent LOS values were correct.

Following this process, the raw booking records were compiled to construct a continuous daily headcount time series. For each continuous calendar day, the number of inmates in the jail was calculated by identifying all the records where the booking date preceded or coincided with that day, and the release date had not yet occurred. This approach effectively

aggregated individual-level intake and release records into a continuous daily census spanning the full range of available dates.

As a result, the Average Daily Population figures were computed on an annual basis for each jurisdiction. A 365-day moving average (MA) was applied to the data to smooth short-term fluctuations and reveal long-term trends in facility utilization. After visual inspection, it was clear that the overall population oscillated between 330 and 500 inmates from 2013 through 2025, with a general uptick in numbers in recent years with the start of the MOUs with Culpeper.

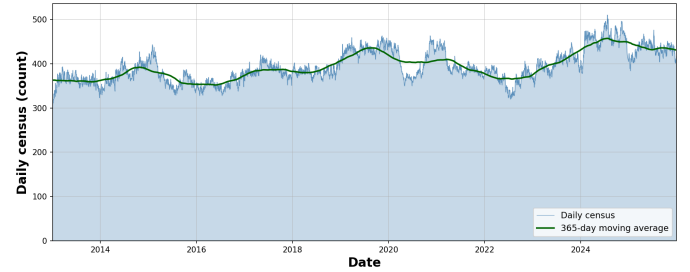


Fig. 1. Daily census of CVRJ inmate population with a centered 365-day moving average.

Fig. 1 depicts the daily jail census at the CVRJ ranging from 2013 to 2025, based on the de-identified dataset. Each day count reflects the total number of inmates in custody on that date, derived by adding an additional inmate based on the booking date and removing an inmate based on the release date. The raw daily series (shown in blue) captures the short-term variation that comes naturally with weekend booking patterns or court-cycle rhythms, while the green line depicts the 365-day moving average to reveal the slower-moving, year-scale trajectory of occupancy. The window is centered, indicating that each smoothed value reflects an equal number of days before and after the focal date, minimizing the short-term deviations. Together, these lines illustrate where the occupancy lies on any given day while displaying the underlying long-run level at each moment.

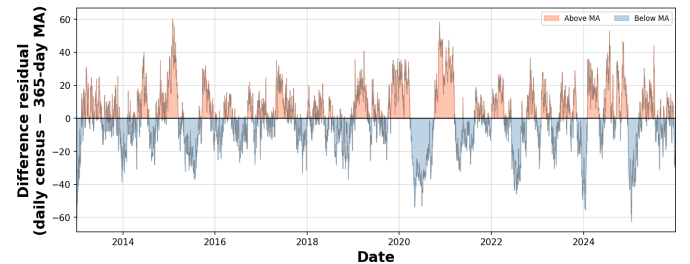


Fig. 2. Residuals plot showing positive values that are the days above the smooth trend; negative values are the days below it.

Fig. 2 plots the difference between the raw daily count and the centered 365-day moving average (difference = daily census – smoothed trend). Days on which occupancy exceeded the year-scale baseline appear as red and positive, while days

below appear as negative blue values. When read in sequence with Fig 1, the two figures separate the trend and residual components: Fig. 1 shows the level and long-run direction of jail occupancy, while Fig. 2 quantifies the operational variability that persists even after the trend is accounted for. The series exhibits persistent patterned oscillation from 2013 to 2015, with difference magnitudes ranging between ± 40 inmates and extreme deviations approaching ± 60 . A notable cluster observed around 2020-2021 is likely attributable to COVID-19 detention policy changes and court system disruptions. For planning purposes, the long-run bed needs and short-run fluctuation place different demands on facility capacity; conflating the two could lead to either over- or under-projection of the required bed needs for CVRJ inmates.

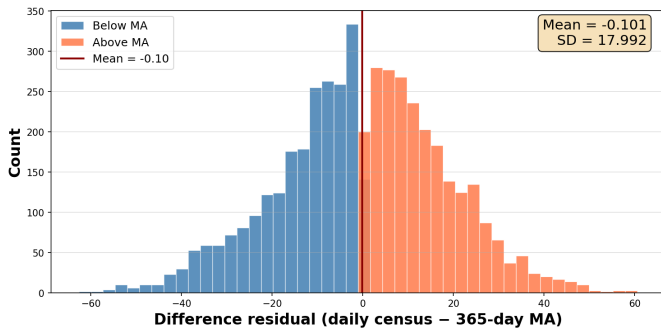


Fig. 3. Distribution of the Difference Residuals

Fig. 3 presents the empirical frequency distribution of the residual series. The year 2020 is excluded from the histogram so that the COVID-19 policy changes do not skew the difference of residual values. The distribution is unimodal and approximately symmetric centered near zero. The mean is 0.126 with a standard deviation of 17.992 inmates per day. The near zero mean confirms that the 365-day MA provides an unbiased representation of the central tendency of the daily census, with no systematic over- or underestimation across the observation period. Most of the residuals fall within the ± 30 range, although the maximum values of the tails extend to approximately ± 65 . These residual diagnostics suggest that the 365-day MA effectively captures the dominant trend, and the short-term variation has an approximately normally distributed noise with a standard deviation of around 18 beds, which is relevant to the capacity threshold planning.

III. MODEL DESIGN

A. Model Objectives

The objective for the model was to provide an estimation of daily jail bed utilization for CVRJ across the next ten years in which Culpeper joins as a member jurisdiction. The model aimed to supplement those projections with quantified probabilities of exceeding the jail's maximum capacity of 660 beds, allowing CVRJ's board to assess the impact of Culpeper's potential inclusion.

B. Modeling Process

The modeling process features a pipeline consisting of four stages:

- 1) *Exogenous Variable Forecasting*: To account for demographic growth within each county region, historical county populations from 2012-2025 were utilized to predict estimates through 2036. This was accomplished by using an ordinary least squares linear regression model. The first-degree polynomial fit model produced a straight line-of-best-fit, establishing continuous and stable population estimates for the forecasted decade. This step was crucial for the ensuing models.
- 2) *Baseline Autoregressive Forecasting (SARIMAX)*: A Seasonal Autoregressive Integrated Moving Average with exogenous variables (SARIMAX) model was implemented to forecast jail inmate populations for the five established CVRJ member counties. The model featured an order of (1, 1, 1). The autoregressive parameter, set at 1, leveraged the temporal momentum from the previous year. The first-order difference, set at 1, transformed non-stationary, trending population data to a stationary series. The moving average parameter, set at 1, corrected for recent predictive variance. Although SARIMAX supports seasonal parameters, the seasonal order was set to (0, 0, 0, 0) after assessing the annual ADP data, which unveiled a lack of statistically meaningful seasonal pattern. Incorporating seasonal terms into the model would create no improvement to the model's fit. While the de-identified CVRJ dataset permitted daily tracking, this baseline SARIMAX model was intentionally constructed using annual ADP to ensure temporal synchronization with the Culpeper County forecast, which was strictly limited to annual data. Moreover, the county population projections were incorporated as exogenous variables, allowing for the forecast to be additionally influenced by general demographic shifts rather than solely extrapolated ADP patterns. This process structure agrees with prior literature applying ARIMA-class models to prison population forecasting [6].
- 3) *Ratio-Based Forecasting for Culpeper County*: Culpeper County's forecasted annual ADP was modeled individually using a mean ADP-to-population ratio forecasting method rather than a time series model. The decision to employ a separate model was determined because of two main constraining factors: the Culpeper ADP data consisted of only 12 annual data points from 2014-2025, and 2024's ADP presented an anomaly with an ADP of 312.33 in comparison to its adjacent years of 200.24 and 227.81. Integrating this susceptible and volatile data into a time series model would be statistically inadvisable and could lead to overfitting and instability issues. Instead, the ratio method calculates the historical mean of (ADP/county population) over 2014-2025 and consequently multiplies the ratio by the Culpeper County population projections of ten years. This more

conservative approach targeted for high-variance data ultimately provides annual ADP forecasts for 2026-2036.

- 4) *Monte Carlo Uncertainty*: A natural limitation of modeling annual ADP is that it fails to consider the daily upswings in fluctuations that may surpass the maximum capacity despite the average remaining underneath. Hence, a Monte Carlo framework was constructed to quantify this potential risk and uncertainty. The process was applied separately for both the CVRJ baseline and Culpeper forecasts. For each forecasted year, a daily arrival rate was calculated by dividing the ADP by the mean LOS. Subsequently, a Poisson distribution with rate parameter $\lambda = ADP/\overline{LOS}$ was used to simulate the number of daily bookings, where \overline{LOS} denotes the mean historical length of stay [9]. For every arrival, the code drew a stay duration from the actual historical LOS data. By stacking the arrivals and overlapping stays, synthetic noise is created. The resulting daily census simulations were transformed into smooth trend lines. The noise series were generated independently for the CVRJ baseline and Culpeper, and the model summed their respective trend lines to produce combined daily load projections.

C. Forecasted Values

The combined CVRJ daily projections were assessed against the maximum bed capacity of 660. The probabilities of CVRJ surpassing the threshold were calculated using the fraction of simulated days within a given year where the census exceeded 660. In the scenario in which CVRJ accommodates Culpeper, the expected probability is 17.6% within five years and 27.2% within ten. In the contrary scenario in which CVRJ remains in its current state without Culpeper, the expected probability is 0.0007% within five years and 0.018% within ten.

TABLE II
JAIL UTILIZATION FORECAST VALUES

Year	Expected Population (Beds)	With Culpeper Prob. Exceeding 660 Beds	Without Culpeper Prob. Exceeding 660 Beds
2031	569	~17.6%	~0.0007%
2035	584	~27.2%	~0.018%

D. Limitations

The main limitation within this study was the absence of daily booking records for Culpeper County. By constraining the data to only 12 annual values containing an anomalous spike, the modeling process was split into two separate modeling pipelines: a SARIMAX model for the existing five CVRJ counties and a simple-ratio forecast for Culpeper County. The annual Culpeper data precluded the opportunity for seasonal factors, as sub-annual patterns are difficult to detect from yearly observations alone. Had the daily booking data been available, a single SARIMAX model would have

been constructed, which would append a seasonal feature, potentially improving confidence through another predictive factor. Another limitation is the model's inability to account for unforeseen circumstances like the probable addition of data centers within Culpeper County, leading to a potential increase in population and thus a corresponding increase in criminal justice demand.

E. Findings

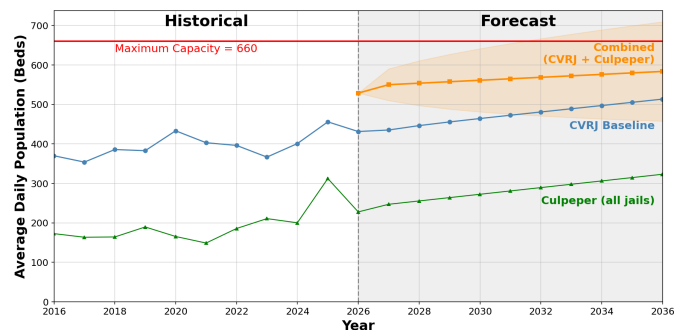


Fig. 4. Jail utilization graph split based on historical values and forecasted projections.

Fig. 4 presents the ten-year forecast of ADP under two scenarios: CVRJ's baseline in its current state (CVRJ + Culpeper MOUs) and CVRJ with the full inclusion of Culpeper County. Under the baseline trajectory, CVRJ's ADP is projected to increase from 430 beds in 2026 to around 515 beds by 2036, remaining well below the facility's maximum capacity. Under the scenario combining Culpeper, there is an increase from 530 beds in 2026 to 590 beds by 2036. The upper bound of the confidence band shows that there is a possibility that the combined scenario exceeds the maximum capacity within the latter years of the ten-year forecasted window, however the most probable outcome suggests that there is adequate capacity to accommodate the combined population.

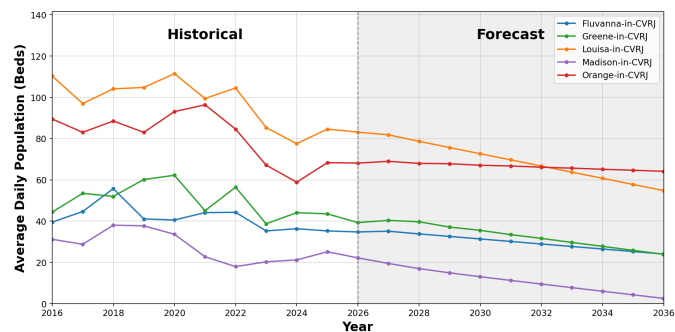


Fig. 5. Jurisdiction-level ADP forecasts

Fig. 5 disaggregates the CVRJ baseline into the individual member jurisdictions within CVRJ to provide a more granular view of the historical ADP trends and ten-year projections. Historically, Louisa and Orange have consistently contributed the most to the ADP value with Louisa peaking near 112 beds

around 2020 before a declining trend through to the mid-2020s. By 2036, the projected ADP for Louisa and Orange converges around 65 beds. Orange, in contrast, shows relative stability across both historical and forecasted periods. The rest of the jurisdictions, Fluvanna, Greene, and Madison, are projected to experience modest declines in their ADP contributions over the forecasted period with Madison trending towards near-zero bed utilization to the CVRJ by the mid-2030s. The collective declining ADP contributions from these counties partially offset the accommodation of Culpeper, supporting the feasibility of such an addition. Moreover, these projections are provided to support member jurisdiction representatives to assess their individual demand trajectories and evaluate their long-term participation within the regional facility framework.

IV. CONCLUSION

We find that Central Virginia Regional Jail can accommodate Culpeper County as a member jurisdiction without exceeding its maximum capacity of 660 beds within the following ten years. Under the potential scenario in which Culpeper is included in CVRJ, the accumulated Average Daily Population is projected to be at approximately 569 beds by 2031 and 584 beds by 2035 with probabilities of exceeding the bed limit at 17.6% and 27.2%, respectively. These probabilities indicate that sustainable accommodation is feasible.

Additionally, it is crucial to take into account that uncertainty compounds as time progresses. The predictions and probabilities for 2036 inherently hold less confidence than those for 2031. It is important to view these projections as more of a planning tool for directional decision-making than as guarantees for the future.

In the future, attaining Culpeper daily booking records would allow for a seasonal factor to influence and strengthen the SARIMAX model, contributing to a more reliable and tighter forecast for Culpeper's projections. Furthermore, a continuous relationship with CVRJ to collect ongoing data would allow for an updated version with adjustments accordingly.

In summary, the findings support CVRJ's capacity to accommodate Culpeper County, while underscoring the importance of further data collection and model refinement to maintain the accuracy of long-term capacity planning.

ACKNOWLEDGMENT

The team would like to express appreciation to all those who have assisted in the ongoing efforts to streamline the creation of the model: The Jefferson Trust; The UVA School of Data Science; The UVA Department of Systems and Information Engineering; and all other members of the Central Virginia Regional Jail Team.

REFERENCES

- [1] Z. Afi et al., "High utilizers of the Albemarle and Charlottesville criminal justice system," in *Proc. 2025 Systems and Information Engineering Design Symposium (SIEDS)*, 2025, pp. 227–232. doi: 10.1109/SIEDS65500.2025.11021179.
- [2] S. Banino et al., "Analyzing candidates for home electronic incarceration on return-to-custody rates for inmates," in *Proc. 2024 Systems and Information Engineering Design Symposium (SIEDS)*, 2024, pp. 505–510. doi: 10.1109/SIEDS61124.2024.10534739.
- [3] Commonwealth of Virginia Compensation Board's Local Inmate Data System, "Average daily population (ADP) for Culpeper inmates," Unpublished dataset.
- [4] F. Dyer, Personal Interview, Jan. 23, 2026.
- [5] J. Lee, "CVRJ bed forecasting," GitHub. [Online]. Available: <https://github.com/jonaslee17/cvrj-bed-forecasting>
- [6] B. S. Lin, D. L. MacKenzie, and T. R. Gulledge, "Using ARIMA models to predict prison populations," *J. Quant. Criminol.*, vol. 2, no. 3, pp. 251–264, 1986. doi: 10.1007/BF01066529.
- [7] T. McDaniel, "De-identified dataset for CVRJ," Unpublished dataset, Central Virginia Regional Jail.
- [8] Virginia State Police, "Virginia crime online," [Database]. [Online]. Available: <https://va.beyond2020.com/>
- [9] S. M. Ross, *Introduction to Probability Models*, 13th ed. Academic Press, 2024.
- [10] New York City Department of Correction and Mayor's Office of Criminal Justice, "Jail Population Forecast Terms and Conditions Report," New York City Council, Mar. 2025. [Online]. Available: <https://council.nyc.gov/budget/wp-content/uploads/sites/54/2025/03/DOC-MOCJ-FY25-TC-Report-Population-Forecast-1st-half.pdf> [Accessed: Apr. 13, 2026].