Predicting Employee Turnover and the

Effect on Staffing of an Organization’s Workforce

Eric D. Alexander

Grand Canyon University

DSC-580: Designing and Creating Data Products

Dr. Michelle Bennett

July 13, 2022

Predicting Employee Turnover and the Effect on Staffing of a Workforce

## Obtaining Data

Data is collected by manual entry of information into a Human Resources Information System (HRIS) by employees or members of the organization (Barisic et al., 2022). This includes Hire or Termination Dates, Job Profile, Compensation, Location, Department, and many more attributes. From the HRIS, reports are built to expose different data regarding the above attributes or other Business Processes including Promotions, Transfers, Pay Raises, and other relevant information pertaining when attributes were modified or changed (Barisic et al., 2022).

These reports then are fed into a Data Warehouse using Python with little transformations applied. Data from HRIS reports are structured to be ready for analysis by the Business Process, Attribute, or Object (Barisic et al., 2022). Each report lives in its own unique table, for example the Business Process table shows when Promotions, Transfers, or Pay Raises took place along with who submitted the request, when it was approved, when it went into effect, and more. This Business Process table can then be joined to the Employee table (Which is fed from a different HRIS report) that contains all most up to date information for each Employee.

## Scrubbing and Cleaning Data

Because data is pulled from the HRIS in a formatted matter, little data manipulation is needed before feeding into a Statistical Model (Barisic et al., 2022). To easily sum up the number of active employees within a time period, another table is built to show a period of dates (I.e., Jan 1st, Feb 1st, Mar 1st or Jan 1st, Jan 8th, Jan 15th, Jan 21st) along with the Employee ID and another two fields titled ‘IsActiveHC’ and ‘IsTerminatedHC’ to indicate whether an individual was actively working within the observed period date. Turnover Rate is then Calculated as the Number of Terminations / Number of Active Employees (Valluru 2018).

Before feeding data into the ARIMA model that is used for Predicting Turnover, exponential smoothing will be applied to create a stationary dataset that is easier to forecast with (Valluru 2018). However, given that exponential smoothing is a technique specific to ARIMA models, a detailed process will be covered in the Exploratory Analysis step.

## Exploring and Visualizing Data

Initial exploration of data starts with understanding the natural Turnover of the company overtime (Peters et al.,1981). A plot of the Turnover Rate for each period being observed can give insight into what this looks like. Before feeding into our model, ARIMA models need stationary data in order to predict with a higher confidence interval. Linear datasets are easier to predict with than Logistic datasets where growth overtime is harder to capture (Valluru, 2018).

Chart, histogram

Description automatically generated

*Note.* An ADF statistic test is used to determine if the dataset is stationary

## Modelling our Data

A variety of algorithms can be applied here to transform the dataset into one that is more stationary. In the below example, the logarithmic value has been taken of the Turnover Rate values followed by a shift differential to observe the change between each Period Date. The results end up as the plot below.

Chart, histogram

Description automatically generated

*Note.* Notice within this figure the ADF statistic number is lower than the prior image, the p-value is below our .05 threshold, and the rolling mean/standard deviation are a flatter and more consistent rate.

## Interpreting our Data

After the stationary dataset is run through an ARIMA model, if exponential smoothing has been applied prior to running the model, the inverse function will need to be applied to convert the output values into actual Turnover values (Valluru 2018). Once actual Turnover values have been obtained, visualize of the final dataset results can be plotted to show the visual pattern of previous patterns and trend to the Forecasted Turnover values.

Chart, scatter chart

Description automatically generated

*Note.* Solid dots are past actual turnover values, while dashed dots are the predicted turnover values

Finalized results should leave little room for interpretation (Lawrence et al., 2021). A confidence interval can be included in the finalized visual to indicate whether the prediction for a specific date is of high or low confidence (Valluru 2018). In the example above, a confidence interval might be wider during the tenth through twentieth week, while a tighter interval will occur during the latter part of the year.

Room for interpretation is left within the action that proceeds delivery of the information (Lawrence et al., 2021). If turnover is being predicted during a period of growth in the company, then results may be actioned on more directly to stay ahead of the potential number of terminations. If risk of the organization being overstaffed is present and no growth is occurring, then leadership may ignore the number of terminations being forecasted as the schedule to backfill the positions may not be as prioritized (Lawrence et al., 2021).

**Pipeline**

An overview of the pipeline from obtaining data to interpreting data can be seen below. All data being stored for the model is recorded at an employee level that is obtained from the HRIS and then stored in a Data Warehouse. Featuring engineering is conducted to build a table to sum up an active number of employees along with the number of terminations. Initial exploratory analysis is done to determine is the dataset is stationary and if not, transformations are applied to do so. The model is than ran, final values are obtained and visualized. The model is then continually monitored and approved upon where it can be.

Diagram

Description automatically generated

## Graphical User Interface

The Graphical User Interface (GUI) provided below is the user facing object that individuals will interact with. Tableau has set functions, so full customization is not available within this GUI, but there is still room to manipulate and change the way data is looked at. Additionally, Tableau utilizes an extract with a scheduled refresh time, so there is no need to load data in from the user side.

Graphical user interface, chart, application

Description automatically generated

*Note.* The darker line is the forecasted values, and the lighter lines are historical values. Hover to see more information including the specific week and forecasted value.

## GUI Features & Elements

Date filters are provided to shorten the range of the Forecast. 52 weeks is the normal default range and can’t go further than this but can be shortened If needed. Date Interval allows you to switch between looking at the turnover and forecast of each week or month. Location allows you to switch between different cities where facilities are located, and Management Level allows you to view Hourly or Professional job titles.

TabPy is not utilized for this dashboard, therefore only major locations that experience high turnover are allowed for Selection in the Location dropdown. The script therefore runs on the backend in the early morning for preset locations and outputs the results into a Database. These database tables are what feeds into the Tableau Dashboard to populate the results. This also means that no data cleaning, processing, or user-end input is required to generate the reports.

Therefore, the main features and elements of this dashboard are the Date Range, Date Interval, Location, and Management Level drop down filters. Below the main visualization there is also a worksheet view that populates. This allows you to view each forecasted value in succession discreetly.

## Pipeline Overview

The pipeline below is executed automatically by means of a virtual server using batch commands. Once data is input into the HRIS system, batch command python scripts run early in the morning and pull fresh data into the SQL Warehouse. This includes the feature engineering that converts Active Headcount Binary table creation. Model exploration is done in model preparation, and once the best model has been selected, a python script is set up again using batch commands to run and output the forecasted values. These values are then fed into another Database table which are used to populate Tableau. Since Tableau is set to run using extracts, fresh data is automatically pulled into the dashboard when relevant. In the end this requires no pipeline execution from the user and the whole pipeline will be carried out in sequence by design.

## Conclusion

The GUI for this Predictive Model is meant to be free of user input with no overly complex means of manipulating data. While only certain individuals have access to the dashboard and have been properly educated on how to interpret results, the dashboard serves as a self-service tool so data can be observed when convenient for Leadership. This way data is easy to consume and digest in the fast paced, agile environment the business operates in.

# References

Barisic, A. F., Tomic, M., & Bach, M. P. (2022). Adoption of Human Resource Information

Systems: Impact of Industry, Size and Market. 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Information, Communication and Electronic Technology (MIPRO), 2022 45th Jubilee International Convention On, 1197–1202. https://doi-org.lopes.idm.oclc.org/10.23919/MIPRO55190.2022.9803480

National Academies of Sciences, Engineering, and Medicine. (2020). Facilities Staffing

Requirements for the Veterans Health Administration—Resource Planning and Methodology for the Future. Washington, DC: The National Academies Press (NAP). <https://doi.org/10.17226/25454>

Peters, L.H., Jackofsky, E.F. and Salter, J.R. (1981), Predicting turnover: A comparison of part-

time and full-time employees*.* J. Organiz. Behav., 2: 89-98.

<https://doi.org/10.1002/job.4030020204>

Valluru, S. (2018). Turnover Analytics & Forecasting Hiring Demand: HR Domain.

International Journal of Business Insights & Transformation, 12(1), 3–5.

Lawrance, N., Petrides, G., & Guerry, M.-A. (2021). Predicting employee absenteeism for cost

effective interventions. Decision Support Systems, 147, N.PAG. <https://doi-org.lopes.idm.oclc.org/10.1016/j.dss.2021.113539>