REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

REVA Academy for Corporate Excellence (RACE)

Pattern Discovery and Forecasting of Attrition

Saumyadip Sarkar
SRN: R19MBA07

Date: 27-Aug-22

MBA in Business Analytics

Capstone Project Presentation
Year: II

race.reva.edu.in

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956, Approved by AICTE, New Delhi

Agenda

Introduction

Background | Current status | Why this study

- High Attrition (~ 34% avg. last 3 years)
- Missed SLAs – 278 in last 14 months
- Penalties - $2,082k for last 14 months

Forecast Attrition in advance → Plan workload management → Reduce Penalties

Transportation Line of Business

Literature Review

Seminal works | Summary | Research Gap

20+ research papers and article are researched to understand employee attrition & timeseries forecasting related work.

Key Findings

Attrition is a very costly affair for any industry. The direct costs of workforce turnover include the cost of hiring new employees, the cost of training new employees, the time it takes to transition, the cost of temporary employees, the cost of lost expertise, and the cost of the job itself (Chakraborty et al., 2021).

In the study (Kumar Jain et al., 123 C.E.) , several machine learning algorithms like Decision Tree, SVM, Random Forests have been used to estimate if an employee will leave or not.

There is one study that stands out from the rest is the use of Ensemble Model Based on Machine Learning Algorithms for automated employee attrition prediction (Alsheref et al., 2022).

In another study (Fallucchi et al., 2020), Gaussian Naïve Bayes classifier has been used to classify if an employee will attrit or not. XGBoost classifier has also been used to classify employee attrition (Jain & Nayyar, 2018).

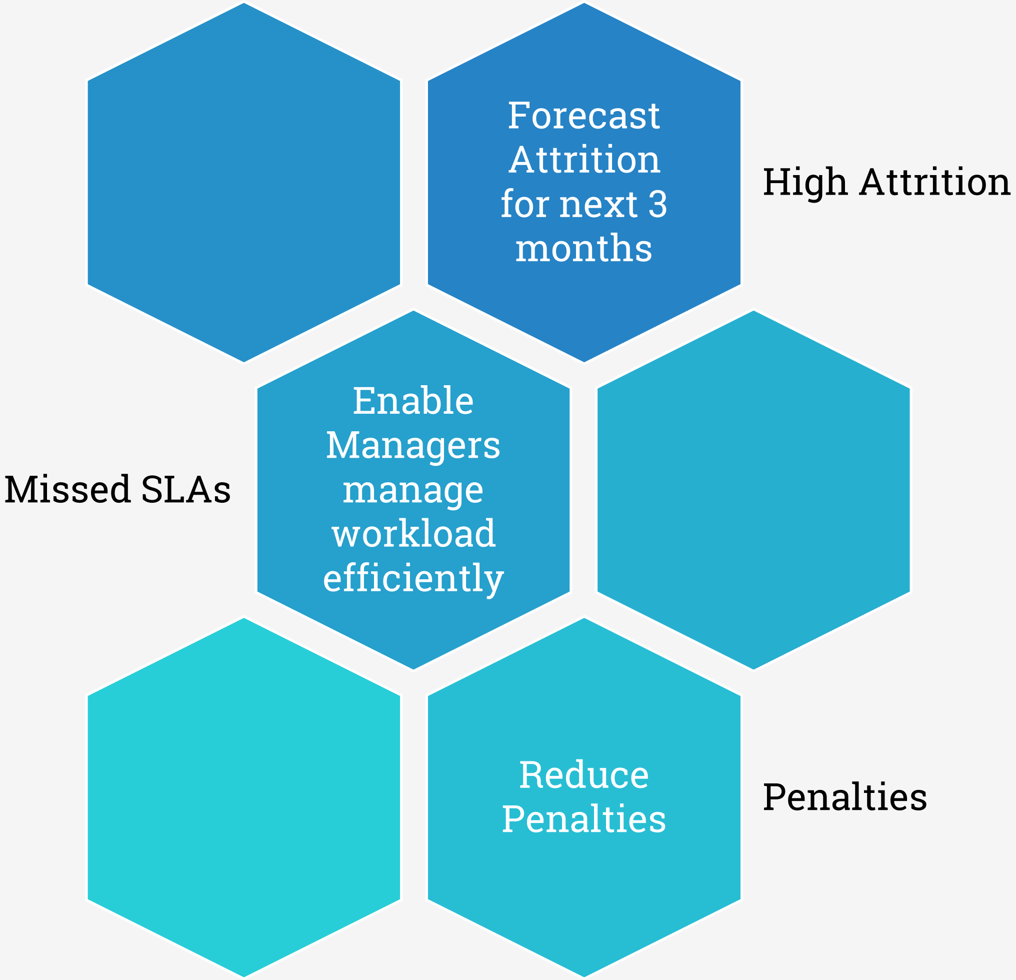
Time series analysis and forecasting are important for a variety of applications, including business, the stock market and exchange, the weather, electricity demand, cost, and usages of goods like fuels and electricity, etc., and in any setting where there are periodic, seasonal variations seen (Mahalakshmi et al., 2016) .

As seen in most of the studies related to attrition predictions, classification is the go-to approach. However, using time-series techniques to forecast future attrition has not been explored enough based on the observations during the literature review. **This establishes a unique opportunity for this study.**

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Problem Statement

Business Problem | Analytics Solution



Project Objectives

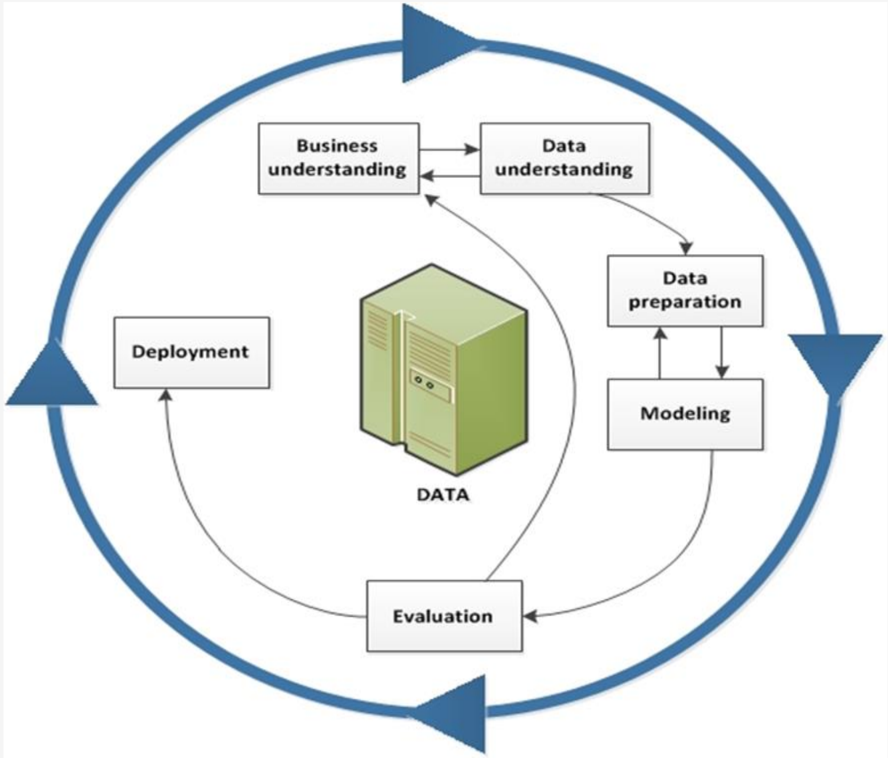
Primary & Secondary Objectives | Expected Outcome

- Identify the required attrition dataset. In this case, only the attrition data of last 14 months are available.
- Study data at various levels and categories to identify trends, patterns, and top contributors which then will be used to create subsets of the main dataset for modeling.
- Explore various timeseries techniques to identify the best timeseries forecasting model which can be used to forecast future attrition.
- Propose early alert system which will help managers manage workload better to reduce missed SLAs and penalties.

Project Methodology

Conceptual Framework | Research Design

CRISP-DM	Business Understanding – Transportation Line of Business
	Data Understanding – Attrition Data of last 14 months
	Data Preparation – Creating subsets by introducing various categories
	Modeling – Explore various timeseries forecasting models
	Evaluation – Compare results of various modeling techniques to identify the best one
	Deployment – Suggest a deployment strategy



Business Understanding

Business Impact | Challenges | Monetary Impact

The transportation one of the most profitable units in this organization.

Solutions	Description
Automated Tolling	Captures vehicle details when a tolling both is crossed and bills customer accordingly. A team also works on dispute resolutions pertaining to technical failure, failed auto-debit attempts, customer complaints, etc.
Automated Parking	It provided intelligent parking solutions mainly for governments. The solution involves fee collections, dynamic pricing, enforcement solutions, etc.
Public Safety	It provides automated photo enforcement, traffic violation solutions, etc.

High Attrition → Missed SLAs → Penalties

(~ 34% avg last 3 years)

278 in last 14 months

$2,082k in last 14 months

Constant change in management | Lack of focus and support

Data Understanding

Data Collection | Variables

Data used in this study is only attrition data collected for the last 14 months. Other data points collected are Number of Missed SLAs & Penalties Paid at contract level for last 14 months.

Fields in the Attrition Dataset

Employee details

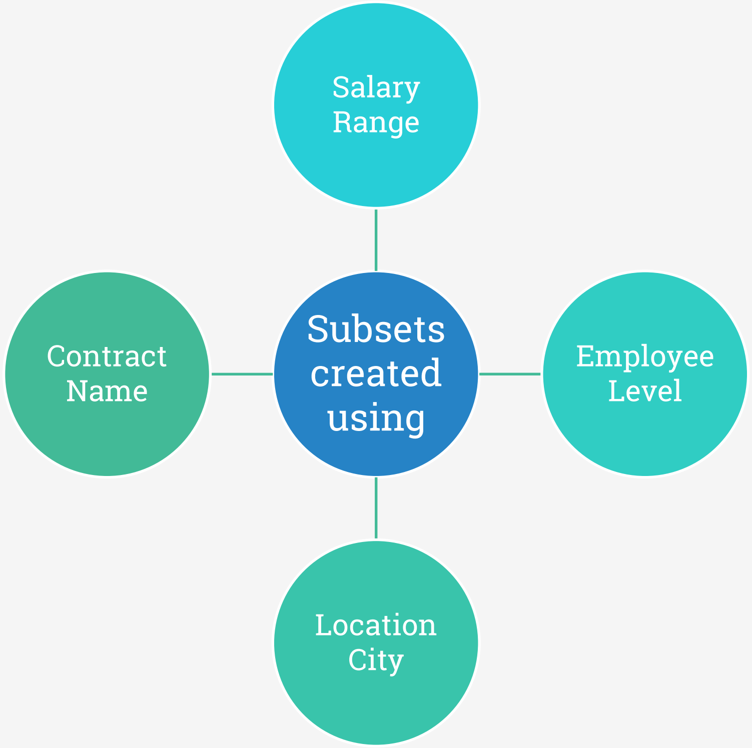
- ID, Employee Name, **Salary**, Last Performance Rating

Employment Details

- Employee Type (Regular or Contract), Joining Date, **Termination Date**, **Employee Level**, Type of Termination, Termination Code, Cost Centre, Job Name

Contract Details

- **Contract Name**, Sector, Business Category, **Location City**, Country



Data Preparation

Pre-processing | Techniques

Based on Termination Date, the data was divided into Quarters and Months.

Subsets created for modeling	
	Overall Attrition by quarter and month-wise
	Attrition by top contracts, quarter and month-wise
	Attrition by top employment levels, quarter and month-wise
	Attrition by top contracts and top employment level, quarter and month-wise
	Attrition by top Cities, quarter and month-wise
	Attrition by top salary ranges, quarter, and month-wise

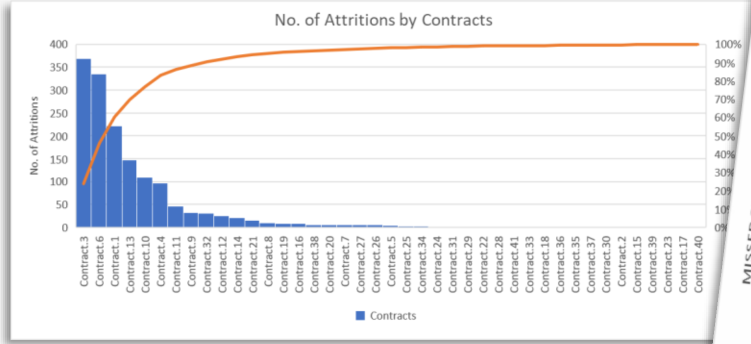
Quarter	Month	Overall	Top 6 Contracts	C01 Employee Level	Top 6 Contracts (C01)	Top Salary Bucket	TSA Top Cities
		Attrition	Attrition	Attrition	Attrition	Attrition	Attrition
Q1- FY22	Month 1	56	46	39	35	46	41
	Month 2	99	85	73	65	79	85
	Month 3	87	73	66	62	72	76
Q2 - FY22	Month 1	100	80	87	71	84	80
	Month 2	106	84	77	68	86	83
	Month 3	95	76	69	61	78	79
Q3 - FY22	Month 1	85	71	67	62	69	71
	Month 2	125	110	107	98	111	108
	Month 3	125	103	93	88	96	102
Q4 - FY22	Month 1	137	106	99	91	100	106
	Month 2	116	103	96	92	102	103
	Month 3	117	100	96	91	101	102

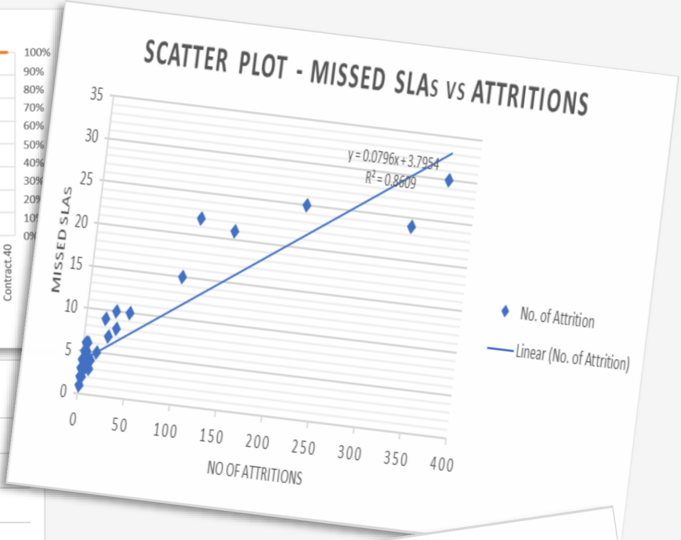
Descriptive Analytics

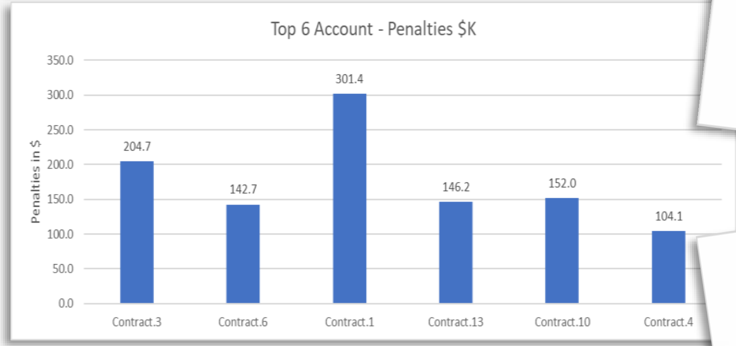
Multivariate Analysis | Hypothesis

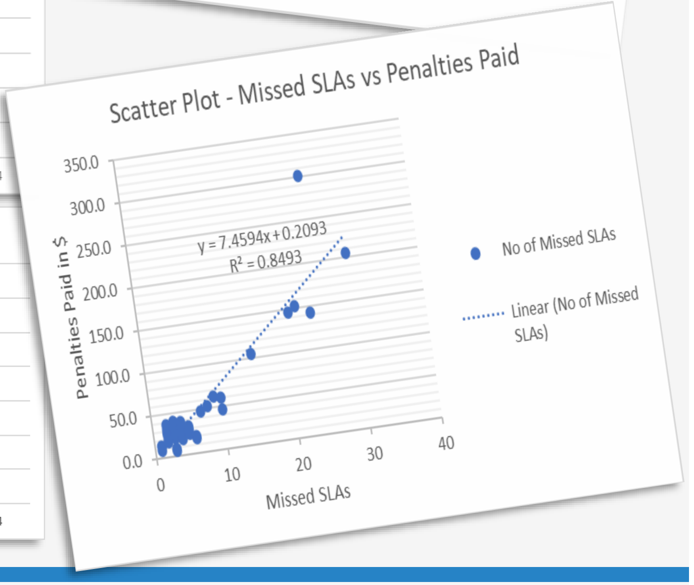
Key Findings

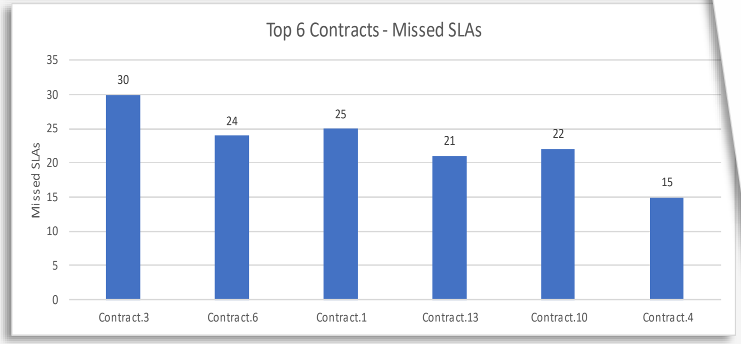
- Top 6 contracts which representing 15% of overall contracts are contributing to more than 80% of attrition and 50% penalties
- ~ 90% of attrition is happening at the junior most level (C01). Top 6 contracts remained constant.
- Top 11 cities representing 27 % contributed to 80% of attritions
- Salary ranging from $20k to $40k contribute to more than 80% attrition
- Strong correlation of 0.93 exists between contract level attrition numbers and missed SLAs
- Strong correlation of 0.92 is observed between Missed SLAs and Penalties paid at contract levels









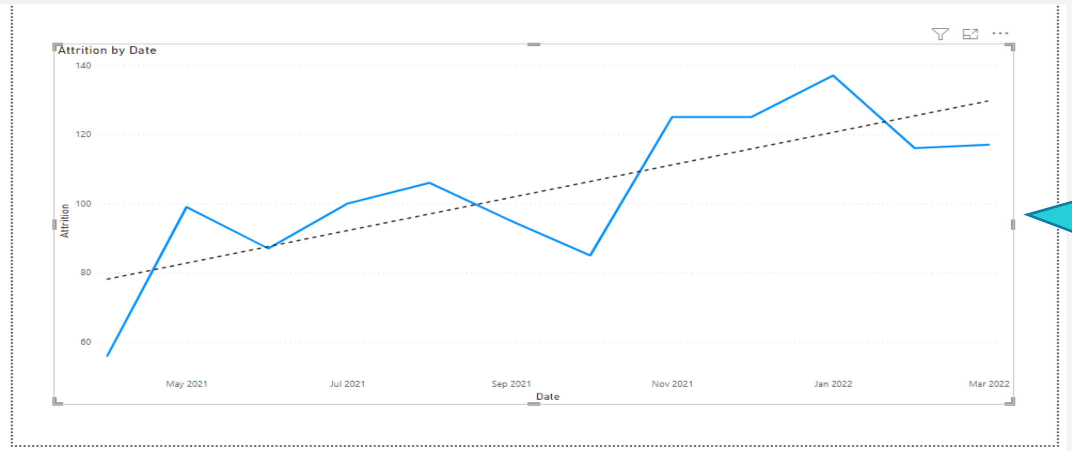


REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956, Approved by AICTE, New Delhi

Modeling

Modeling Techniques | Modeling Process | Model Building

Dickey-Fuller Test → Data is Stationary Across all Dataset



Upward trend across all datasets contradicts Dickey-Fuller test

Following timeseries model are tested to validate both Stationary & Trend in the datasets

Modeling Technique Used using MS Excel	Moving Average
	Ratio to Moving Average
	Exponential Smoothing

Modeling Technique Used using Python	ARIMA
	LSTM
	Holt-Winters

Modeling

Modeling Techniques | Modeling Process | Model Building

Moving Average

- A 3 months moving average is considered as the data is divided into quarters consisting of 3 months
- It is an empirical methods for smoothing and forecasting time-series

Ratio to Moving Average

- Building on Moving Average, some additional steps are added like deseasonalizing the data, building regression model on the deseasonalized data to forecast
- This method is frequently used to show the data's overall movement without taking seasonal effects into account.

Exponential Smoothing

- MS Excel's FORECAST.ETS function is used
- It uses AAA version of Exponential Smoothing
- This method is based on three smoothing equations: stationary component, trend, and seasonal.

Modeling

Modeling Techniques | Modeling Process | Model Building

ARIMA

Auto ARIMA is used to find the best combination of the order (p,d,q) → The best order found was (1,0,0). → This order was used for all other datasets

LSTM

The data is divided into train and test data and MinMax preprocessing technique is used on both the datasets → LSTM learns using "TimeSeriesGenerator" function with input as 2. → The Sequential, Dense, and LSTM classes are called from the Keras library → The model is the run 50 epoch to understand the stage at which the loss is lowest for it to learn on its own

Holt-Winters

Single HWES is used as the first technique. → Double HWES is tried next with both additive and multiplicative for trend → Finally Triple HWES is used with both additive and multiplicative for trend and seasonality

Modeling

Modeling Techniques | Modeling Process | Model Building

Linear Regression

Given forecasted attrition, missed SLAs are predicted using a linear regression model.	Similarly, with forecasted missed SLAs, the penalties are predicted using a linear regression model.

Regression Formulae
Predicted Missed SLAs = (0.07958 * Forecasted Attrition) + 3.7954
Predicted Penalties = (7.45942 * Predicted Missed SLAs) + 0.20933

Model Evaluation

Results | Interpretation | Insights

ETS or Moving Average has the best MAPE across all the datasets.

TS Models	MAPE					
	Overall Data	Top 6 Contracts	C01 Employee Level	Top 6 Contracts at (C01)	Top Salary Range	Top Cities
Moving Average (3)	9%	6%	7%	7%	7%	6%
Ratio to Moving Average	12%	11%	13%	12%	9%	12%
Exponential Triple Smoothing (ETS)	10%	6%	4%	4%	2%	4%
ARIMA	14%	14%	15%	19%	13%	15%
Holt Winters ES1	25%	26%	28%	29%	25%	29%
Holt Winters ES2_ADD	13%	12%	13%	12%	19%	13%
Holt Winters ES2_MUL	18%	25%	15%	25%	23%	27%
Holt Winters ES3_ADD	16%	14%	19%	16%	15%	15%
Holt Winters ES3_MUL	20%	17%	24%	19%	18%	18%

* LSTM has been used only on overall data where the accuracy was 86%.

Model Deployment

Demonstration

The models used here need to be tested with future attrition data to establish the consistency of results. Since other external factors are not considered which can affect attrition, this study would be an ongoing activity. However, the final findings will be shared to gain overall feedback from the management. Based on the feedback a deployment process can be decided.

Proposed Solution –
Early Alert System in Power BI

Results and Insights

Key Findings | Suggestions

The forecasting techniques were tested on the actual attrition data of the following two months.

Model Performance on Test Data

Dataset	Forecasting Technique	MAD	RMSE	MAPE	MAPE during training
Overall Data	Moving Average	25.8	28.6	17%	9%
	ETS	14.0	17.5	9%	10%
Top 6 Contracts	Moving Average	6.3	8.6	5%	6%
	ETS	7.1	7.6	6%	6%
C01 Employee Level	ETS	7.7	9.1	6%	4%
Top 6 contracts at (C01)	ETS	7.5	7.5	7%	4%
Top Salary Bucket	ETS	4.6	4.8	4%	2%
Top Cities	ETS	4.1	5.0	3%	4%

ETS is consistent during training & testing phase

Results and Insights

Key Findings | Suggestions

Forecasted attrition numbers are used for predicting Missed SLAs and Predicted Missed SLAs are used for predicting Penalties.

Regression Formulae
Predicted Missed SLAs = (0.07958 * Forecasted Attrition) + 3.7954
Predicted Penalties = (7.45942 * Predicted Missed SLAs) + 0.20933

Overall Data			
Month	ETS (Forecast)	Predicted Missed SLAs	Predicted Penalties in ($ k)
Month 1	128.7	14	104.9
Month 2	133.4	14	107.7
Month 3	138.2	15	110.6
Total		**43**	**323.2**

Regression Statistics (Y = Missed SLAs, X = Attrition)	
Multiple R	0.93
R Square	0.86
Adjusted R Square	0.86
Standard Error	2.75

Regression Statistics (Y = Penalties, X = Missed SLAs)	
Multiple R	0.92
R Square	0.85
Adjusted R Square	0.85
Standard Error	23.18

Conclusion and Future Work

Proposed solutions | Scope for future work

Given only attrition data, future attrition can be predicted with greater accuracy.

FORECAST.ETS
(MS Excel
Function)

- Consistent across all datasets
- Simple & Easy to Use
- Greater acceptability within organization

As with any modeling technique, continuous training with future attrition is recommended.

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956, Approved by AICTE, New Delhi

References

Bibliography | Webliography

Chakraborty, R., Mridha, K., Nath Shaw, R., & Ghosh, A. (2021). Study and Prediction Analysis of the Employee Turnover using Machine Learning Approaches; Study and Prediction Analysis of the Employee Turnover using Machine Learning Approaches. *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*. https://doi.org/10.1109/GUCON50781.2021.9573759

Alsheref, F. K., Fattoh, I. E., & M.Ead, W. (2022). Automated Prediction of Employee Attrition Using Ensemble Model Based on Machine Learning Algorithms. *Computational Intelligence and Neuroscience*, *2022*, 1–9. https://doi.org/10.1155/2022/7728668

Fallucchi, F., Coladangelo, M., Giuliano, R., & de Luca, E. W. (2020). *Predicting Employee Attrition Using Machine Learning Techniques*. https://doi.org/10.3390/computers9040086

Mahalakshmi, G., Sridevi, S., & Rajaram, S. (2016). *A survey on forecasting of time series data; A survey on forecasting of time series data*. https://doi.org/10.1109/ICCTIDE.2016.7725358

Kumar Jain, P., Jain, M., & Pamula, R. (123 C.E.). *Explaining and predicting employees' attrition: a machine learning approach*. https://doi.org/10.1007/s42452-020-2519-4

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956, Approved by AICTE, New Delhi

Annexure

Additional Information | Plagiarism score

Pattern Discovery & Forecasting of Attrition for Transportation

ORIGINALITY REPORT

6%	3%	0%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	global.oup.com Internet Source	2%
2	Submitted to Visvesvaraya Technological University, Belagavi Student Paper	1%
3	Submitted to ILSC - Sydney Student Paper	<1%
4	Submitted to University of Glasgow Student Paper	<1%
5	Submitted to Coventry University Student Paper	<1%
6	Submitted to University of London External System Student Paper	<1%
7	Submitted to University of Strathclyde Student Paper	<1%
8	ieeexplore.ieee.org Internet Source	<1%
9	gupea.ub.gu.se Internet Source	<1%
10	library.itc.utwente.nl Internet Source	<1%
11	www.vodppl.upm.edu.my Internet Source	<1%
12	minerva-access.unimelb.edu.au Internet Source	<1%
13	www.scinapse.io Internet Source	<1%

Exclude quotes On
Exclude bibliography On
Exclude matches < 10 words

REVA UNIVERSITY
Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Thank you!

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Annexure

Publications | Conferences