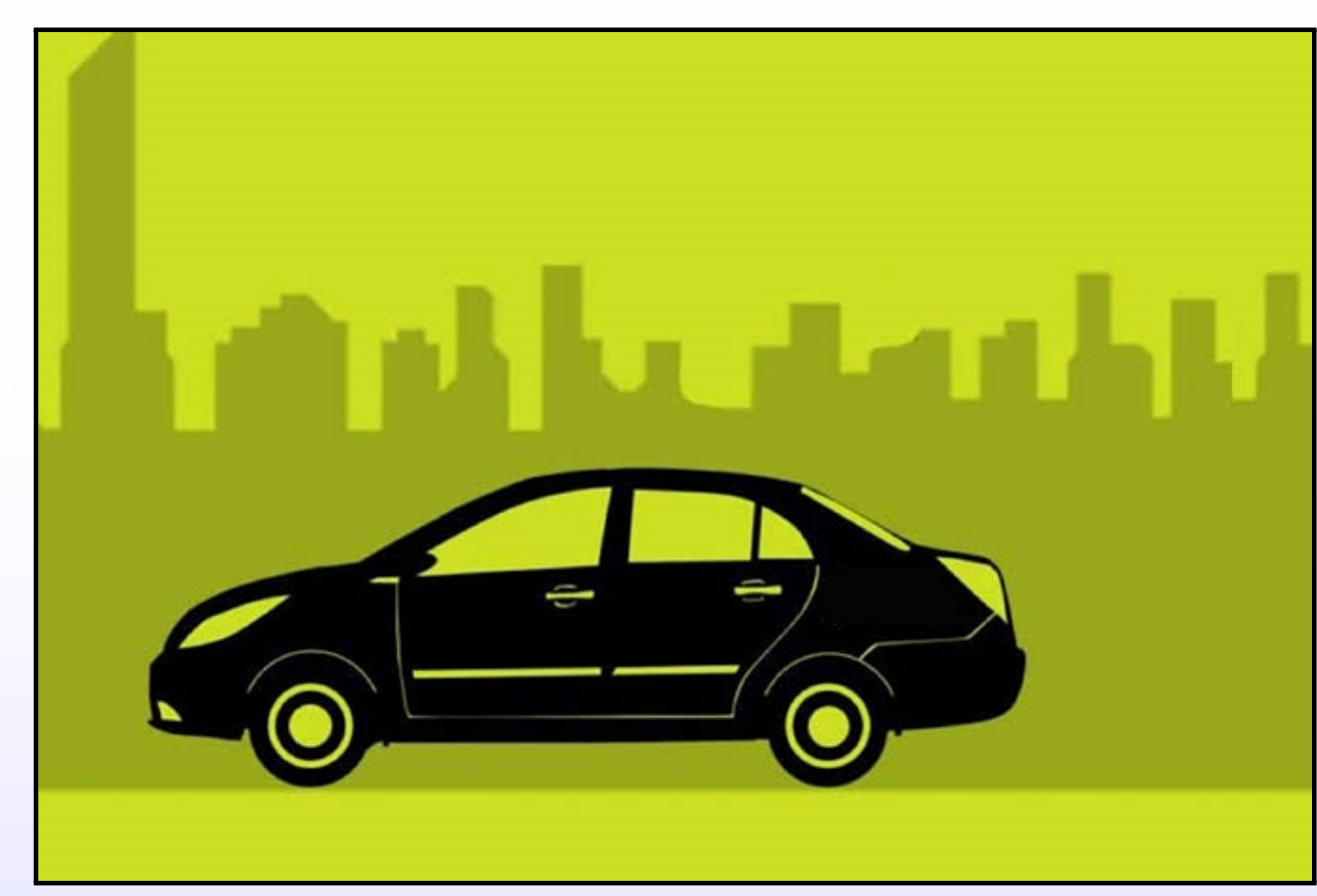


# Trends in OLA Cab Rates: A Conformal Prediction Analysis

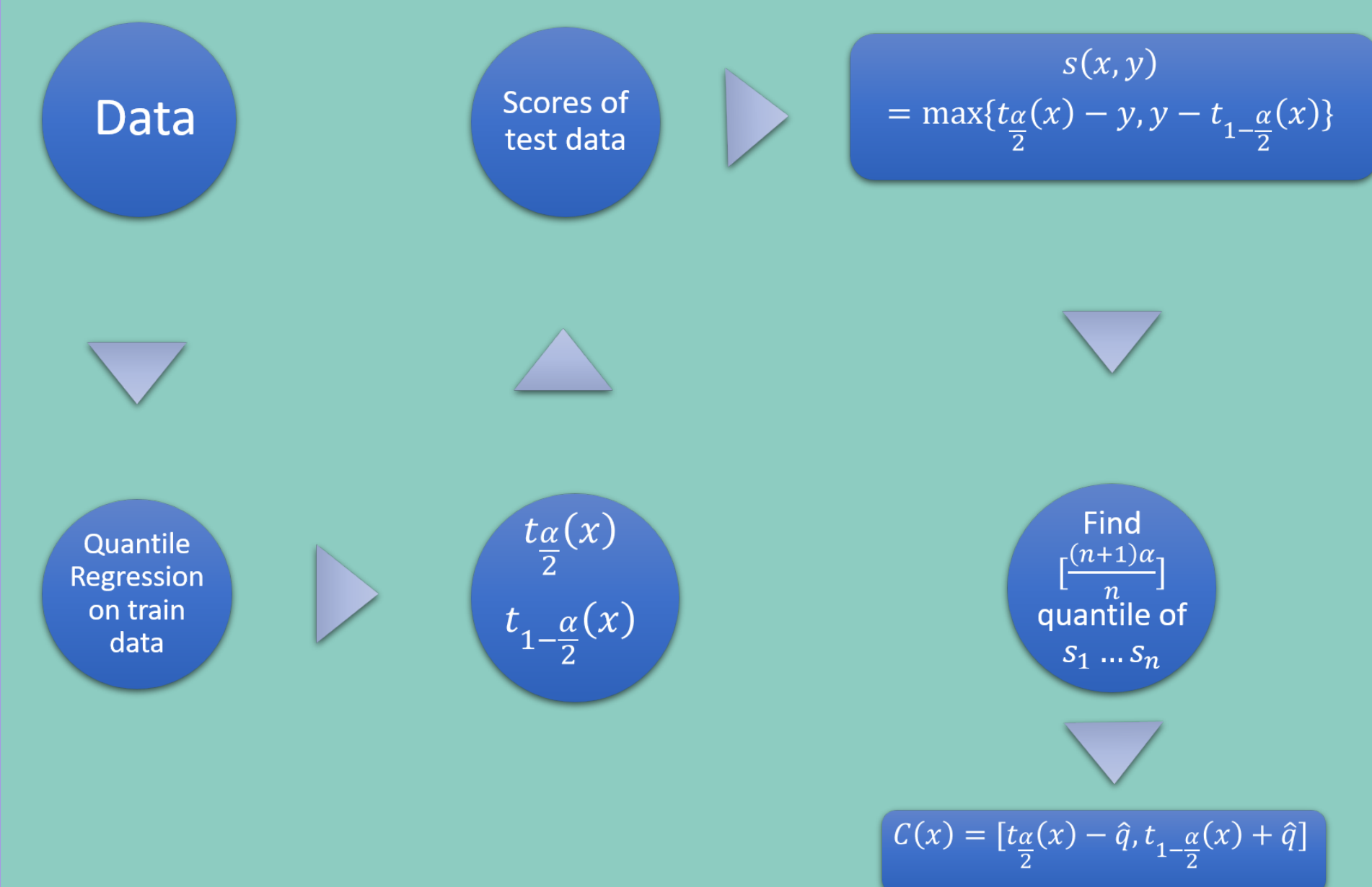


## Objective

The aim of this work is to gain insights into the functioning of Ola cab service with the help of data [1]. Our objective is to recognize trends and patterns within the data, such as the most popular time of day for riders, the most commonly used locations and the types of services in demand. Additionally, given a new data point, we seek to make a prediction of the total trip cost based on conformalized quantile regression. Moreover, we plan to identify potential areas of improvement, growth and any issues that may need to be addressed with the service.

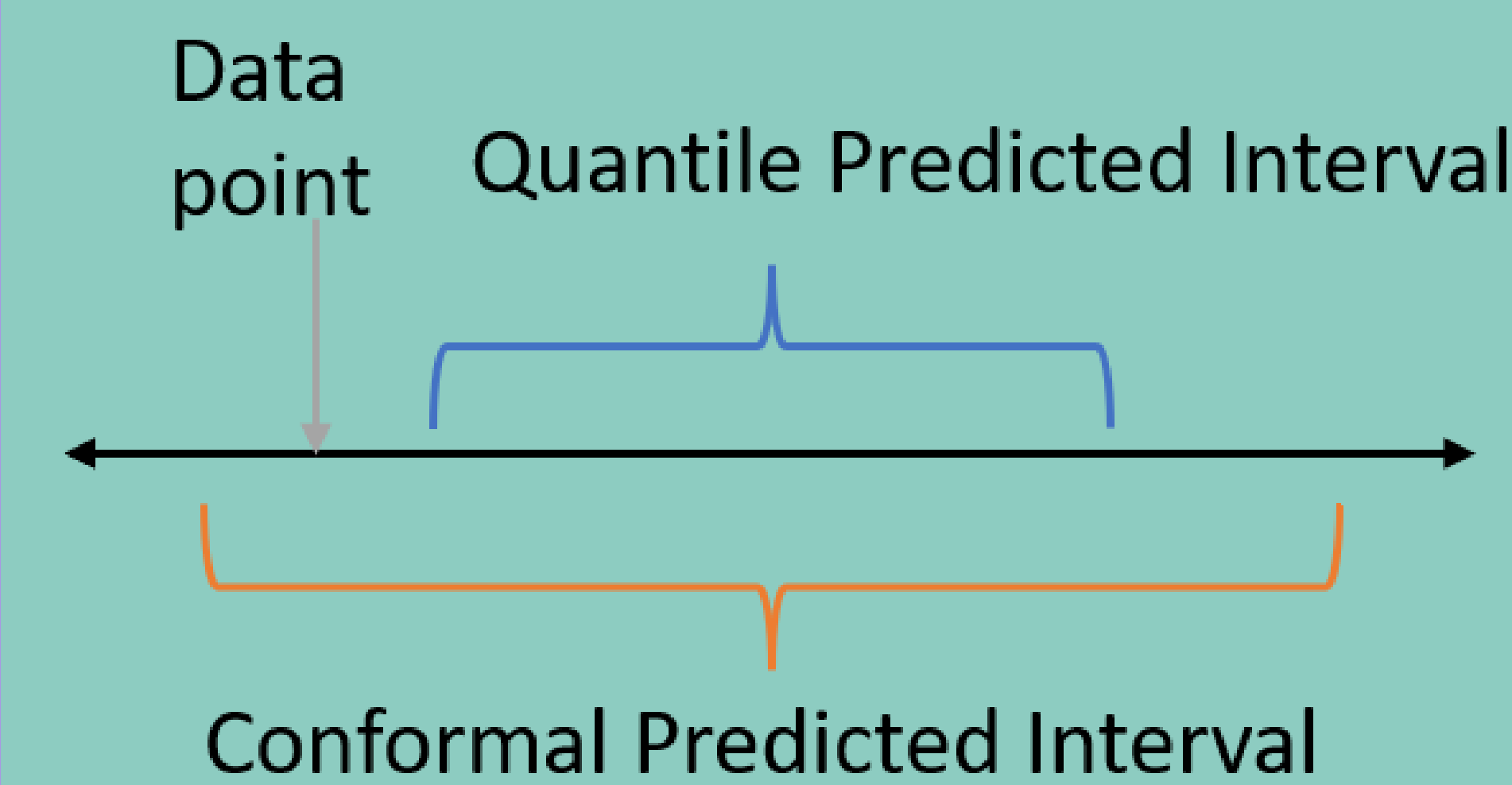
## Models

We used **conformal quantile regression**[2, 3] to create prediction intervals for a set of calibration data. The flowchart demonstrates the procedure:



## Example

Here we show a visualisation of the conformal quantile regression. We adjust the quantiles by the constant  $\hat{q}$ , picked during the calibration step.



## References

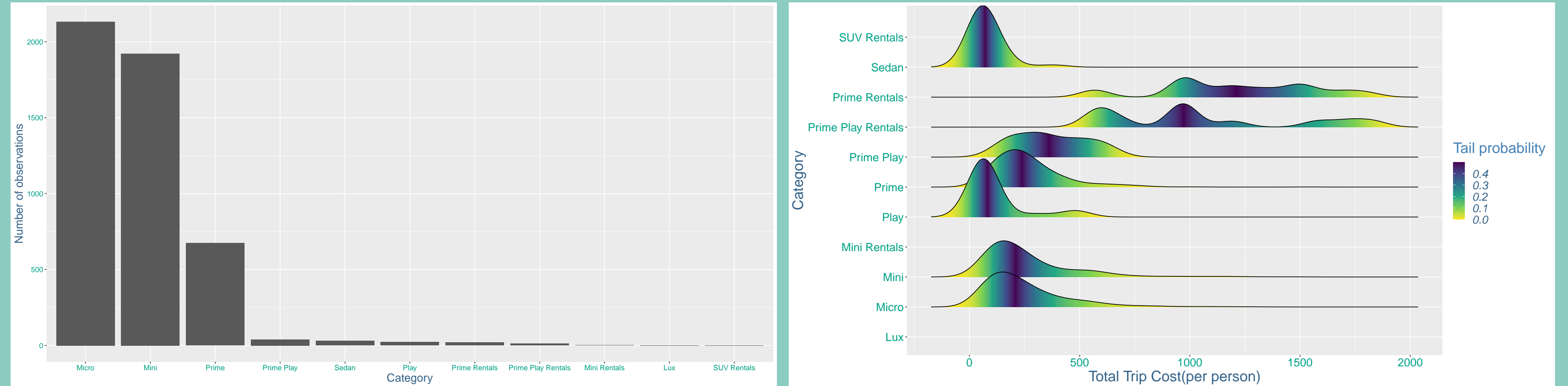
- [1] OLA Trips. (2022, September 1). Kaggle. <https://www.kaggle.com/datasets/nimish23/ola-trips>
- [2] Angelopoulos, A. N., & Bates, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*.
- [3] Shafer, G., & Vovk, V. (2008). A Tutorial on Conformal Prediction. *Journal of Machine Learning Research*, 9(3).

## Acknowledgements

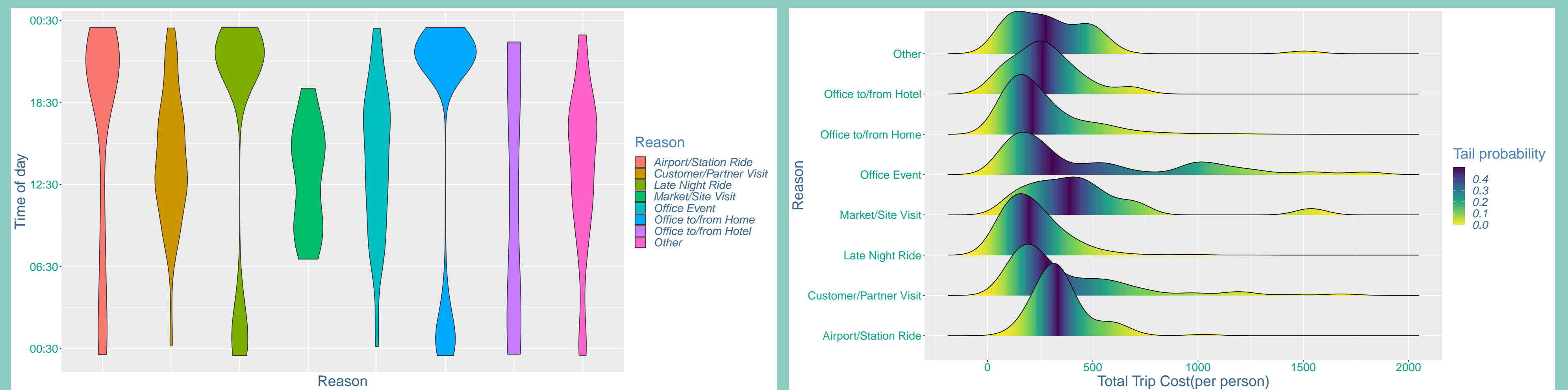
We are thankful to the International Statistical Literacy Project for providing us with the opportunity to present our poster.

## Preferences are determined by cost-effectiveness and necessity

The plots demonstrate that, on average, Micro, Mini, Play and Prime are cheaper and more widely used than Prime Rentals and Prime Play Rentals, with the total trip cost being evenly spread. However, Sedan has low costs but is less used. This may be due to its comparatively bigger size; people usually do not travel in groups using Ola Cabs.

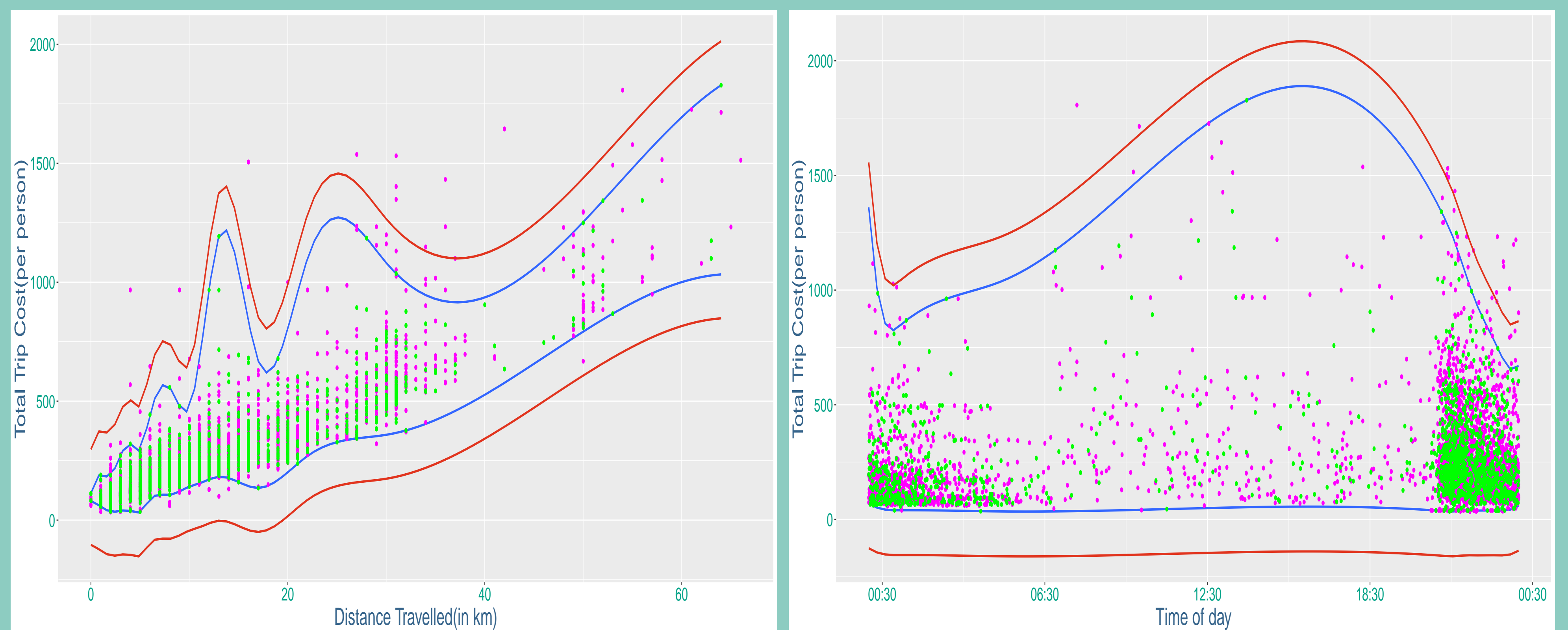


Cabs are typically used during times of commuting from home to office, or vice versa. The total trip cost remains relatively consistent regardless of the reason for travel (but involves distance travelled), with the exception of travelling for office events or market site visits.



## Conformal quantile regression

We trained a quantile regression model on training data, which produced the blue lines to show the upper  $(1 - \alpha/2)$  and lower quantiles  $(\alpha/2)$  for the plots, where  $\alpha = 0.005$ . Most of the green points fell within the boundaries of the upper and lower quantiles, whereas the pink points did not. To improve the prediction intervals, we applied conformal quantile regression to capture most of the pink points.



The usefulness of conformal quantile regression lies in the **efficiency** of the model, which is the probability that a test point lies in the predicted interval. Although quantile regression may estimate the quantiles with high probability but conformalized model provides the exact efficiency of  $1 - \alpha$ . The quantile regression had a accuracy of around 97% which got increased to 99.52% upon use of conformal procedures.

