UNTAPPING THE POTENTIAL OF CROWDSOURCING APP DATA TO MEASURE URBAN BICYCLE FLOW

CURRENT ISSUE

Sweden lacks research when it comes to untapping the potential of bicycle possibilities using crowdsourced data, there are a couple of scientific papers regarding urban bicycle volume, however this doesn't apply to smaller cities such as Lund. A whole lot of research has been done on bigger cities and countries but not as much on smaller municipalities.

Method

The method employed to perform this task was done by gathering volunteers who were willing to let their movements be tracked in real time by installing an application that uses GPS, the TravelVu app. Different roadway links across Lund were picked in order to count cyclists at multiple intersections, with the total being 180.

The data collection was gathered during a certain period, stretching from 7am to 5pm with four 15 minute sessions spread over the day. The data provided multiple properties, including but not limited to travel time, travel distance, trip segments and so on.

A simplified linear model equation was used in order to comprehend the linear relationship between the output and input variables. To identify influential data points Cook's distance formula was used to easier recognize data points that might negatively affect the regression model. Leverage was also taken into account since it refers to the influence a data point has on the estimation of the regression coefficients.

Conclusions

Multiple reasons could have caused these outliers to occur in the first place but it generally indicates error while collecting data, measurement variability or unusual phenomena. The main conclusion was that mix traffic is one of the most influential factors when it comes to inaccurate results, all kinds of motor vehicles affect the precision when gathering data. Especially when solely counting bicycles.

Additionally, obstruction of some sort is also a reasonable source as to why there are so many errors, disturbing the equipment in any way can cause it to get heavily damaged, natural causes like the weather also messes with the installation and even non-recognizable vehicles are influential.

Result

A total of 13 outliers were identified when analyzing the data, they stood out from the rest of the observations which affected the accuracy of conclusions. This is clearly noted as they deviate quite a lot from the norm, especially data points that stray away from the typical pattern or distribution of the rest of the data.

A deeper examination was performed to get a better insight into why certain data points had extremely high or low values, multiple factors were considered to have been the cause for these errors. Some examples are related to old road infrastructures and diverse mix of daily traffic.

The results also showed that these outliers mainly occurred close to the heart of the city, where it's heavily crowded no matter the time of the day, as well as the outskirts of the city, such as roads nearby E22 or at locations where the intersections were located near roundabouts.

Future research

This study strives to give the reader a better understanding of why such inaccuracies occur in the first place and what the cause potentially could be, basically how the quality and quantity of crowdsourced data affect the reliability of predictions for bicycle volume in different urban environments.

It also contributes with valuable knowledge for different purposes such as when modeling traffic safety measures in order to better perceive an accurate bike volume since it's one of the biggest factors related to traffic accidents, which in its turn helps produce better accident prediction models.

Furthermore, it contributes to future research and offers the possibility to flesh it out even further since there's always something new to learn and technology keeps improving day by day.

Sebastian Palomino Achulli