ESSnet Trusted Smart Statistics – Web Intelligence Network

Grant Agreement Number: 101035829 — 2020-PL-SmartStat

Work Package 3

New Use-cases

Deliverable 3.12: Report on assessment of challenges and opportunities

UC6 Faster Economic Indicators using new data sources

Version, 2024-12-18

Prepared by:

UC 6 coordinator(s):
Remy Kamali – Statistics Sweden; remy.kamali@scb.se

Contributors:
Remy Kamali – Statistics Sweden; remy.kamali@scb.se
Kira Gylling – Statistics Sweden; kira.gylling@scb.se
Olav ten Bosch - Statistics Netherlands; o.tenbosch@cbs.nl
Ian Grimstead - Office for National Statistics
Pieter Vlag – Statistics Sweden

This document was funded by the European Union.

Web Intelligence Network

Funded by the European Union

Contents

1 Introduction

This document is part of the Work package 3 (WP3) New use-cases from the ESSnet Trusted Smart Statistics – Web Intelligence Network project (TSS-WIN). The overall objective of WP3 is to explore the potential of new types of web data sources for official statistics. This document highlights UC6: Faster Economic Indicators using new data sources.

This use case aimed to explore publicly available traffic camera data to understand the "busyness" of various towns and cities in the UK by examining street scenes, identifying vehicles and pedestrians, and generating time series of busyness. The focus was on measuring "busyness" as a proxy for the health of local economies, particularly during the COVID-19 pandemic, rather than on traffic flow.

Traffic cameras have proven to be valuable sources of data for understanding activity levels in cities and towns, providing insights into changes in mobility patterns. The UK has a vast network of publicly accessible traffic cameras, managed by national agencies and local authorities. These low-resolution images, while not allowing for individual identification, offer important insights into urban activity. The UK's Office for National Statistics (ONS) recognized the potential of this data and launched a research project in 2020 named the "Experimental Faster Indicator – Traffic Activity" to develop a busyness index for pedestrians and vehicles based on traffic camera data.

In 2021, the ONS partnered with Statistics Sweden within the framework of Use Case 6 to explore and extend the usefulness of this data source across Europe.

2 UK Case Study: Traffic Activity Indicator

2.1 Objectives

The ONS Data Science Campus initiated a research project in 2020 with the following objectives:

- to produce indicators of busyness for urban areas.
- to leverage underutilized public infrastructure (traffic cameras).
- to provide faster insights into economically important traffic disruptions.

2.2 The data source

The UK is equipped with thousands of traffic CCTV cameras, predominantly in cities and towns. London alone hosts over 1,000 cameras with public access. Openly available camera footage can be accessed from locations such as Aberdeen, Glasgow, Kent, London, and Manchester, among others.

From these, a subset of cameras was selected to analysed historical activity trends during the COVID-19 pandemic, spanning urban and rural environments across the UK. Key locations include Durham, London, Manchester, and Northern Ireland, with data collection starting between March and May 2020. These sites represent a broad geographic spread and varied settlement sizes[1].

Camera selection criteria

The study prioritized monitoring **changes in busyness** - the activity level of places - over traditional traffic flow. This included:

- Trends in vehicle and pedestrian counts.
- Cameras showing areas likely to capture pedestrian activity (e.g., sidewalks or areas with visible public movement).

Cameras on trunk roads or rural roads without pedestrian zones were excluded, as they did not accurately reflect overall population busyness.

The study leveraged the UK's extensive CCTV network to measure how activity levels in public spaces evolved during the COVID-19 pandemic, providing a comprehensive view of activity trends during a transformative period, showcasing the versatility of CCTV in understanding social and mobility patterns.

Annotated data for model training and evaluation

For effective training and evaluation of machine learning models, accurate annotation is essential. Using the VGG Image Annotator (VIA) hosted on the UN Global Platform, approximately 4,000 images from London and the North East were labelled across seven categories: **car, van, truck, bus, person, cyclist, and motorcyclist**. This was achieved through volunteer effort:

- Two groups from the Office for National Statistics (ONS) labelled the images. One group labelled images randomly, while the other considered specific time periods to account for varying illumination levels.
- Bias minimization: Random selection and stratified sampling ensured minimal bias and improved model performance evaluation.

[1] For extensive coverage of the study and the topic, see [1, 2]

2.3 Processing pipeline: Cloud-Driven Architecture

To manage large-scale image data and provide continuous uptime, the processing pipeline was built on Google Cloud Platform (GCP), leveraging the flexibility and scalability of cloud technology.

The Data Science Campus at ONS identified key advantages of Cloud Integration:

- Automated maintenance: reduces operational burden.
- Enhanced security: particularly useful for remote working scenarios.
- Scalable storage: accommodates vast datasets effortlessly.
- On-Demand Compute: enables cost-efficient, large-batch processing and model testing.

Pipeline workflow:

- Stateless Cloud Functions: key to robustness, these can scale automatically under heavy demand without risk of corruption.
- Efficient Scheduling: GCP Scheduler triggers processing every 10 minutes or daily, initiating tasks through Pub/Sub Topics.
- Cost Optimization: Charges are based on usage (compute time and memory), enabling rapid scaling with minimal costs.

Performance metrics:

- Over 1,000 images can be downloaded in under a minute, scaling easily to tens of thousands.
- Images are stored in secure "buckets," ensuring shared access for authorized users.
- Processed images generate object counts, which are stored in BigQuery for seamless integration with tools like Data Studio for visualization.

The overall architecture is presented in Figure 1:

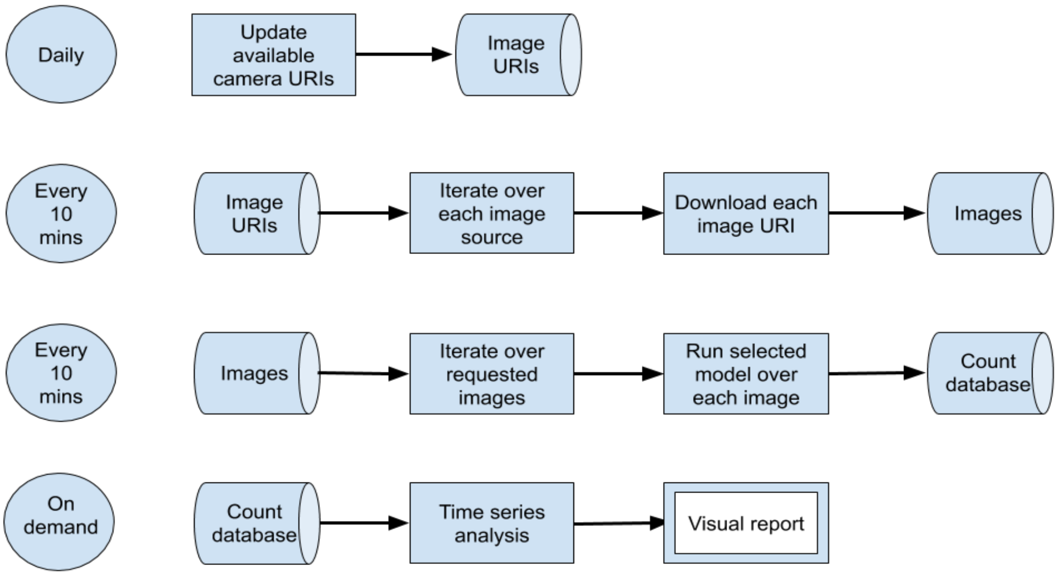

Figure 1: High-level architecture of the system
Source: Office for National Statistics, UK

Time series analysis
Processed data was analysed weekly using a virtual machine on GCP's Compute Engine. This compute-intensive phase required hours to handle imputation but ensured accurate insights into trends.

Broader and flexible applications
The pipeline design is adaptable to various data types beyond imagery. It can process audio files for signal detection or text documents for thematic analysis, highlighting its versatility for diverse analytical needs.

2.4 Deep Learning Model: An Overview

The deep learning model integrates spatial and temporal data to process images efficiently, with distinct workflows for single-image processing (spatial) and sequences of images (temporal). This approach improves object detection, static/moving classification, and data quality through cleaning and validation.

Figure 2 below describes the modelling process. The data flow shown in green processes spatial information from a single image. That is, each image is processed independently without considering other images captured before or after it. The blue flow processes spatial and temporal information for an image and its most recent neighbouring images.

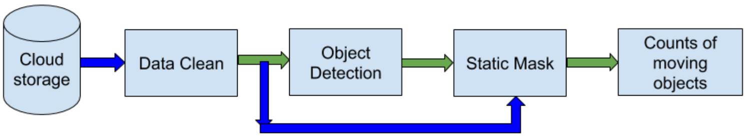

Figure 2: Data flow of the modelling part
Source: Office for National Statistics, UK.

Data cleaning

Data cleaning ensured high-quality statistical outputs by addressing common issues with faulty images, such as:

- **Large monochromatic areas**: Subsampling and colour-level detection are used to identify artificial images with significant single-color coverage.
- **Repetitive rows**: Pixel sequence analysis detects and flags faulty images with repeated rows caused by compression or physical camera issues.

Object detection

The process involved identifying and classifying objects (e.g., pedestrians, vehicles) in images. Several models were tested, including:

- **Faster-RCNN**: Selected for its speed, cost efficiency, and superior performance, achieving higher mean average precision (mAP) and better handling of small objects compared to alternatives like the Google Vision API.
- **Future Exploration**: ONS continues its research about advanced models like YOLOv5 and ensemble methods.

Static mask filtering

Static masks were employed to distinguish moving objects from static ones using structural similarity (SSIM):

- SSIM analyses patterns in images sampled at 10-minute intervals to extract background and filter out static objects, improving the accuracy of activity detection.
- Benefits: The static mask also reduces false positives, such as misidentified objects (e.g., rubbish bins flagged as pedestrians).

Model validation

Model performance was validated against external data sources:

- **Traffic counts comparison**: Results from Faster-RCNN were compared with Automatic Number Plate Recognition (ANPR) traffic counts. While discrepancies occur (e.g., vehicles turning off before the second ANPR camera), short ANPR segments with overlapping camera data provided robust validation.

1. **Spatial and Temporal integration**: As the ONS combined spatial and temporal information, the model adapted dynamically to challenges such as faulty data and varied object movement.

2. **Efficiency and Scalability**: Faster-RCNN demonstrated cost-effective precision, and static masking offered additional refinement by eliminating irrelevant data.
3. **Cross-validation**: Cross-referencing model outputs with ANPR data ensured reliability, even accounting for edge cases like route divergence.

2.5 Statistics/indicators

The essence of traffic camera data analysis lied in transforming raw, incomplete, and seasonally fluctuating data into robust and actionable insights. ONS Data Science Campus achieved this through systematic imputation of missing values, seasonal adjustments, and advanced statistical techniques.

Challenges of raw data:

- Raw traffic camera data often have gaps due to downtime or maintenance, and systematic variations tied to the calendar, making them unsuitable for direct analysis.
- Seasonal adjustments were essential for uncovering trends and enabling comparisons across time periods and locations.

Methodology for processing Data:

- Imputation: Missing data was addressed using the imputeTS package, which leverages seasonally decomposed imputation. This ensured smoother and more reliable time series, with other imputation methods proving less effective or computationally unfeasible.
- Seasonal Adjustments: The ARIMA-based TRAMO/SEATS method adjusted for seasonal fluctuations (e.g., rush hours or weekly cycles). This process decomposes data into:
  - Trend: Long-term progressions.
  - Seasonal components: Recurring patterns.
  - Irregular components: Residual noise after removing trends and seasonality.
  - Adjustments were made for rare occurrences (e.g., negative seasonal values) using transformations, with a square root transformation yielding optimal results.

Faster indicators in production:

- Traffic camera data provide high-frequency (hourly) and timely (weekly) statistics, offering insights into population "busyness" in urban centers.
- These data complement traditional mobility data (e.g., Google Mobility, Highways England), particularly for their granularity and immediate availability.
- These indicators are currently considered as **official statistics in development**. They were previously named experimental statistics.
- The results are handed over to the output group at ONS using an automated data transport schedule.
- The indicators are published every Thursday on the ONS website: [Traffic camera activity - Office for National Statistics (Institutes.gov.uk)]
- The code is publicly available on GitHub: https://github.com/datasciencecampus/chrono_lens

2.6 Measuring "Busyness": A complex issue to tackle

One of the major challenges in this project appeared to be to determine a statistical and conceptual framework for defining "busyness" and to decide whether it is best anchored in a finite population or a random process. This is a multifaceted issue touching on critical aspects of statistical analysis and data interpretation. In the earlier work there was it was not completely clear how this issue was addressed. The available documentation and the published paper do not provide definitive answers. Therefore, below, we outline a structured view on these issues and formulate the beginning of a framework for addressing them in structured way.

Definition of "Busyness"

Conceptual Definition

ONS articulated "busyness" as a measure of activity density, represented by the relative index of footfall and vehicle movement. Future projects at ONS and across Europe should strive to clarify its purpose: Is it meant to reflect real-time dynamics, long-term trends, or specific behaviours in different areas?

Operational Definition

In the studies, ONS specified how "busyness" was derived from the data, relying on proxies such as traffic cameras. However, they acknowledged the absence of attributes like travel purpose or individual demographic details.

Population Framework

We emphasize two primary approaches to defining the population framework:

1. **Finite Population Approach**

- **Definition:** The population could be defined as "all pedestrians and vehicles observed within the camera's field of view over a specific time frame." Some cameras may zoom in/out or change views from time to time. It is hard though to define the camera field/view coverage.
- **Statistical Units:** These units are the observed individuals or vehicles.

Challenges:

- High uncertainty exists regarding whether the captured data (traffic camera images) fully represents the population of interest (e.g., residents, workers, visitors). Some further investigation should be carried out to study a relationship between the busyness and the traditional population data (such as data from Census). This will help us to better understand whether and how the busyness links to the local population and demographic characters.
- Defining boundaries for the population: Should they be temporal (daily, weekly) or spatial (within the camera's range)?

Recommendation: If this approach is chosen for future projects, we recommend explicitly defining the scope and limitations of the finite population. Additionally, acknowledge that this represents a sample or proxy of a larger conceptual population.

2. **Random Process Approach**

- **Definition:** “Busyness” could be treated as a stochastic process where footfall and vehicle movements are random events occurring over time and space. Although ONS did not explicitly adopt this approach, it aligns more closely with their description of “busyness.”

- **Statistical Units:** Events (e.g., individual vehicle or pedestrian detections) occurring according to a random process.

Advantages:

- Focuses on probabilistic modelling rather than strict enumeration, as highlighted in the Board review.

- Avoids the need to define a “complete population.”

Recommendation: This approach is recommended for dynamic, real-time indicators, emphasizing variability and probabilistic interpretation.

Future Work: Evaluating and Discussing the implications

We believe that future projects should encourage detailed discussions on the implications of choosing either the finite population or random process approach. Additionally, limitations related to data sources – for instance traffic cameras, should be addressed, particularly their inability to capture travel purpose or demographic attributes. The suitability of “busyness” for specific use cases (e.g., urban planning, event monitoring) should also be more precisely highlighted.

We suggest that the following elements should be included in future projects:

- A clear definition of “busyness” and the chosen conceptual statistical framework (finite population or random process).

- Acknowledgment of limitations in data collection, particularly the lack of travel purpose information.

- Justification of the chosen approach based on the study’s objectives.

- Suggestions for improvements, such as integrating other data sources (e.g., surveys, geolocation data) to provide richer contextual information.

Concluding remarks

The concept of “busyness,” as used in this case study by ONS, serves as an index reflecting the relative density of pedestrian and vehicle activity, derived from traffic camera data. Given the absence of information about travel purpose or demographic attributes, the study cannot directly enumerate a specific population in the traditional statistical sense.

To address this, we propose treating “busyness” under a **random process** framework, where individual detections represent events occurring within a stochastic model. This approach allows a focus on patterns and trends without needing to define a complete finite population. However, for certain applications (e.g., policy decisions), it may be necessary to approximate a finite population based on observed data and assumptions.

We recommend including a detailed discussion on these points in future projects. This should be accompanied by an acknowledgment of the methodology's limitations and potential enhancements, such as integrating auxiliary data sources to refine the understanding of "busyness".

3 Use Case 6: Expanding the approach across Europe

The goal of Use Case 6 was to extend the UK's traffic camera data approach to other European countries, starting with Sweden. Statistics Sweden became the first external user of ONS's cloud-shared code, running the system on a Windows laptop.

3.1 The Project Objectives

The main objectives were:

- to produce a proof of concept for busyness indicators in urban areas.
- to assess the usefulness of these indicators for economic activity, especially during the COVID-19 pandemic.
- to increase public sector capability to leverage traffic camera data for new indicators.

3.2 Key Activities

The first two years (2021-2022) of the project were the most productive, with key milestones including:

Assessment of Potential Data Sources: TrafikVerket and Webbkameror.se in Sweden

We evaluated two potential data sources for the trial: Trafikverket (the Swedish Transport Administration) and Webbkameror.se. Trafikverket's cameras, being government-owned, were selected for the trial as there were no issues with using a dataset provided by the Swedish government. Additionally, these cameras were comparable to those used in the UK. Conversely, the camera views from Webbkameror.se were less suitable, as they were primarily located at temporary construction sites.

Ultimately, we chose two cameras from Webbkameror.se and 23 from Trafikverket. Of the Webbkameror.se cameras, one was in Stockholm and the other in Kiruna, the latter being selected to assess the effects of snow and darkness on the estimates. All cameras from Trafikverket were located in Stockholm.

Selection of Relevant Cameras

From the approximately 240 cameras provided by Trafikverket in Sweden, the majority were positioned along high-speed roads outside Stockholm (dual carriageways and motorways without pedestrian pathways). Since these locations are not suitable for pedestrian activity, they would have skewed any pedestrian count data. Therefore, a subset of cameras was selected, focusing on urban areas in Stockholm, specifically residential and commercial zones. Ultimately, 25 cameras from these urban areas were chosen for data collection.

Choice of Technology and Setup of Working Environment

At the time, the existing system at the Office for National Statistics (ONS) utilized the Google Cloud Platform (GCP), leveraging its scalability and low-maintenance features in line with a cloud-native design.

However, this setup required potential users to establish a GCP project, which involved accessing paid-tier services and posed a barrier to entry. Additionally, securing the project demanded specialized knowledge.

To address these challenges, the ONS Data Science Campus aimed to develop a version of the system that could be deployed on a single machine with minimal effort and dependencies. The goal was to create a solution that could run on stand-alone Linux, macOS, and Windows computers (ensuring platform agnosticism) and require only moderate computing power—suitable for a standard laptop.

Python was chosen for its cross-platform capabilities, and the system was designed to output CSV files daily, avoiding reliance on any specific database infrastructure. CSV files were preferred for their simplicity, low dependency requirements, and ease of integration with databases. The only key requirement was that the machine needed to run continuously with an internet connection, such as home broadband.

The localhost version of ChronoLens was developed at the Data Science Campus in 2021 and later tested by Statistics Sweden on a local Windows laptop. This solution proved adequate for running the system on a stand-alone machine. However, its main strength is also its primary drawback: all models come pre-trained and cannot be retrained on a typical laptop due to hardware limitations. Consequently, the system is prone to bias, as all training data originated from cameras in the UK. This could be problematic in Sweden, where conditions are darker and snowier for much of the year, potentially affecting the system's performance under these unexpected conditions.

3.3 Proof of Concept Results (Sweden)

System Testing and Results

The system was tested using 23 cameras located in Stockholm and Kiruna. Data collected over several weeks revealed clear activity patterns, including rush-hour peaks for buses and cars, and increased pedestrian traffic during weekends. Figure 3 displays aggregated data from two weeks, with one graph per vehicle type and one curve for each weekday. Below are key observations for each graph, presented from left to right and top to bottom. These results closely mirrored the patterns observed in the UK.

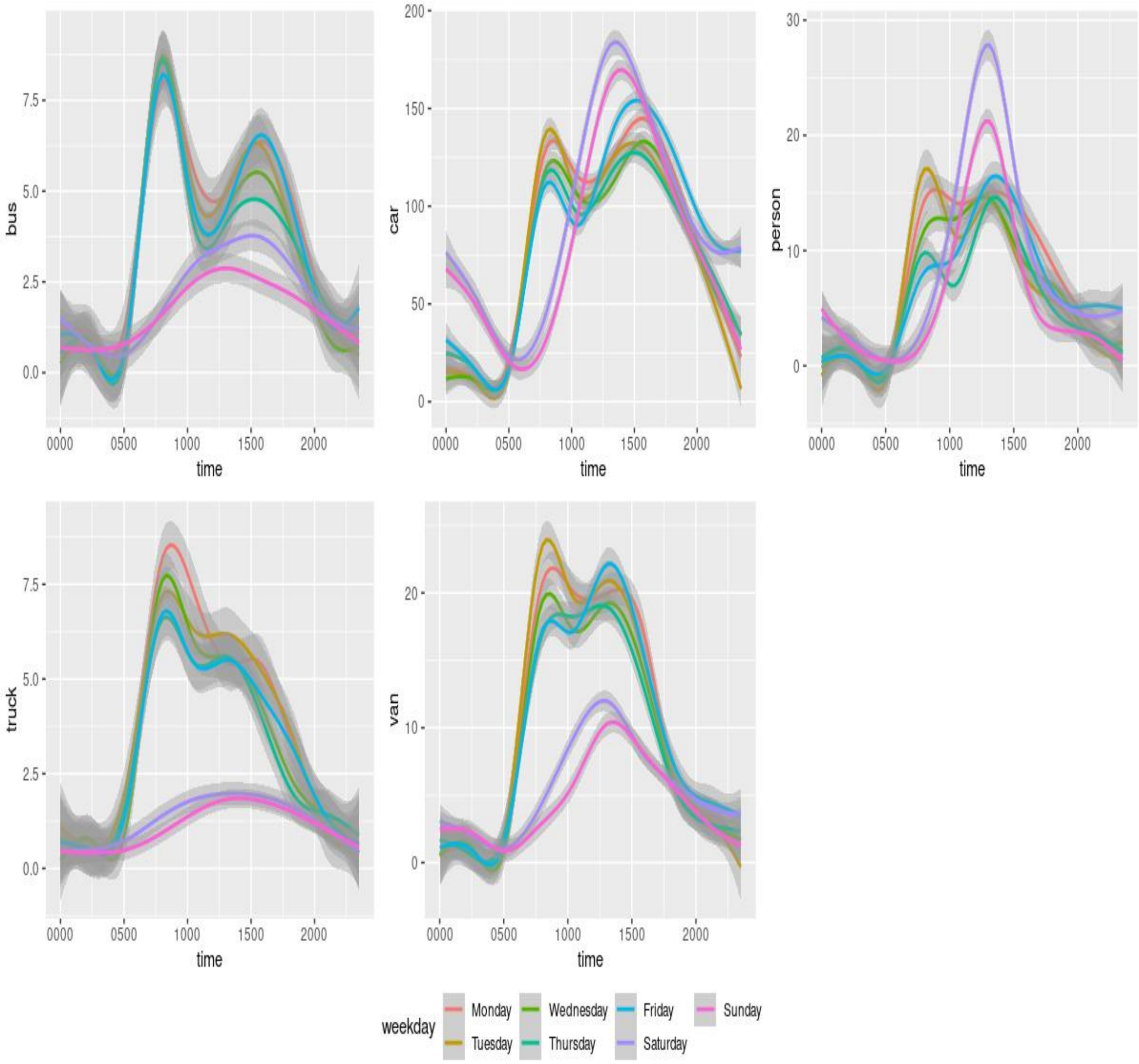

Figure 3. Aggregated data over two weeks

Key Observations from Graphs (Left to Right, Top to Bottom)

1. **Buses:** Peak activity occurs during weekday rush hours—both mornings and afternoons. Weekend activity is significantly lower, consistent with typical bus schedules in Sweden.

2. **Cars:** Similar rush-hour peaks are observed for cars, though the afternoon peak is slightly lower than the morning peak, reflecting daily commute patterns. On weekends, however, there is a notable afternoon peak.

3. **Pedestrians:** Activity remains relatively steady during typical office and store hours, with a distinct afternoon peak on Saturdays and a slightly smaller one on Sundays. This may correlate with the car traffic peaks, suggesting some pedestrians could be weekend shoppers or people heading downtown for lunch.

4. **Trucks:** There is a clear morning peak on weekdays, followed by a gradual decline in activity throughout the day. This pattern likely reflects delivery operations to stores and businesses,

which are predominantly conducted in the mornings. Truck activity is minimal on weekends, aligning with reduced operations during that time.

5. **Vans:** Vans show similar patterns to cars, though weekend activity is lower, and afternoon peaks occur earlier, around midday. The reason for this earlier peak remains unclear.

Figure 4 below displays two plots showing number of faulty (red), missing (green), and running (blue) cameras per day over a week per plot.

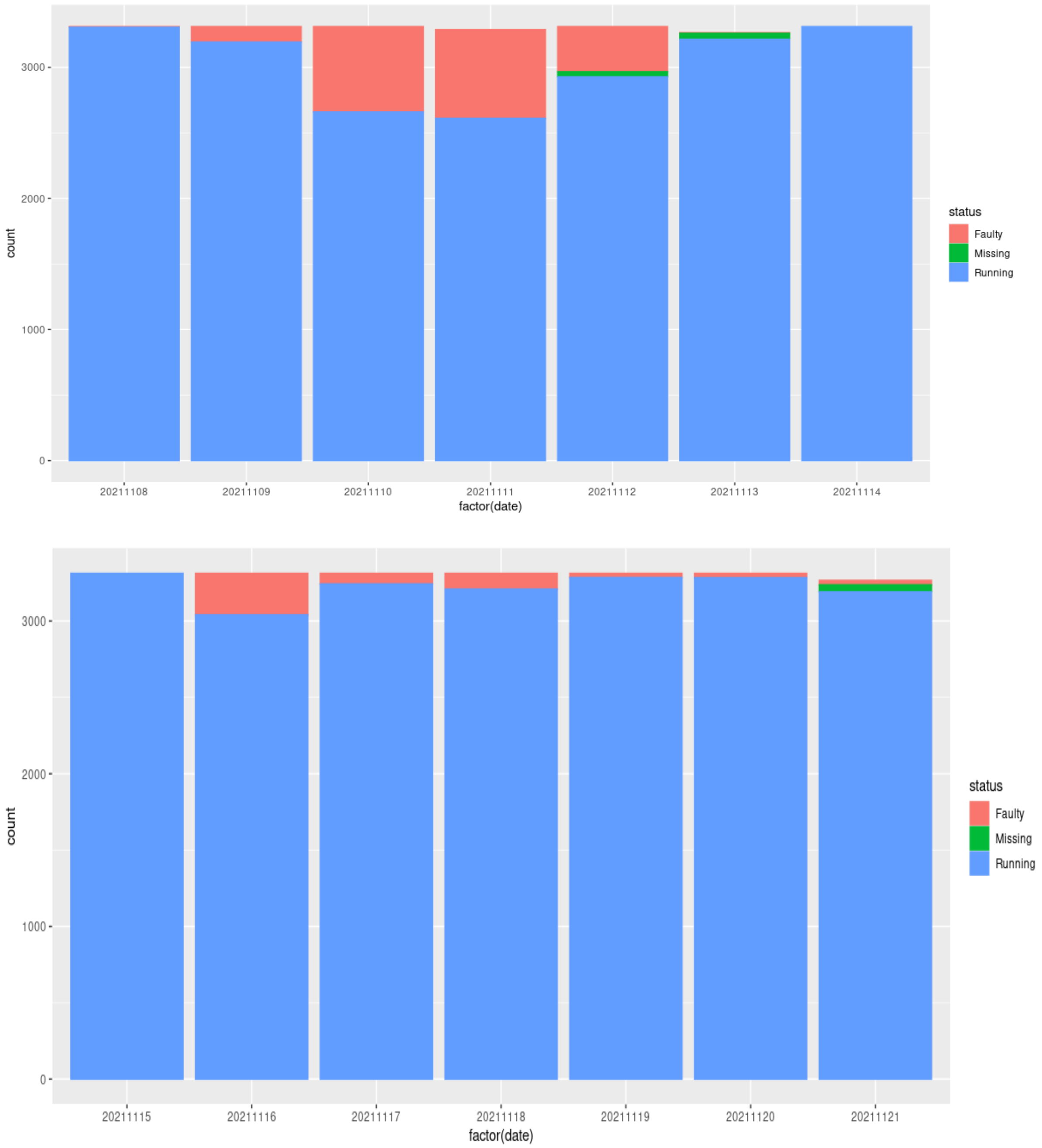

Figure 4. Number of faulty cameras (red), missing (green), running (blue) per day for a week

Despite the reassuring results of the proof of concept, the following challenges were identified:

- Limited access to traffic cameras in Sweden.
- Accuracy issues due to external factor such as weather conditions, lighting, and camera placement.
- The need for additional data sources to complement traffic camera data for a more robust analysis.

4 Challenges, Limitations, and opportunities

4.1 Infrastructure Challenges

Statistics Sweden faced significant infrastructure challenges during this project. At the time, there was no dedicated production environment for Python, necessitating the use of a private laptop for system operations. This setup inherently limited the scale and continuity of data collection efforts. Additionally, the absence of a viable Google Cloud Platform (GCP) solution posed further constraints, a situation unlikely to change in the foreseeable future due to organizational policies and external factors.

However, since the project's inception in 2021, Statistics Sweden has made notable progress in modernizing its technological ecosystem. Efforts are underway to develop a Python production environment, complemented by initiatives to establish a more advanced computing center. Plans also include the creation of an on-premises cloud platform tailored for statistical analysis and machine learning, leveraging open-source tools. These developments mark a significant step forward in building a robust and modern production environment, both in terms of hardware and software. While the applicability of these advancements to analysing traffic data images remains uncertain, this use case underscores critical considerations for the modernization of statistical production processes.

Operating the system on a local laptop introduced its own set of technical hurdles, such as ensuring the machine remained continuously operational and addressing resource constraints that hindered parallel tasks. These limitations highlighted the pressing need for scalable, reliable, and dedicated infrastructure to support complex statistical workflows.

Barriers to Cloud Adoption

Specific barriers to leveraging cloud services at Statistics Sweden include regulatory restrictions, stringent data privacy requirements, and budgetary limitations. These challenges underscore the broader difficulty of adopting cloud-based solutions in public-sector statistical agencies, where operational transparency and data security are paramount.

The Importance of Scalable Infrastructure

A key takeaway from this project is the necessity of scalable infrastructure for conducting high-volume, complex statistical analyses. The current setup at Statistics Sweden, while improving, remains inadequate for meeting the growing demands of official statistics production. Without significant investments in infrastructure, such as high-performance computing and cloud-based platforms, the ability to process and analysed large-scale datasets will remain constrained.

Broader Implications for the European Statistical System (ESS)

These challenges are not unique to Statistics Sweden but resonate across the European Statistical System (ESS). Addressing them at the ESS level offers an opportunity to drive systemic improvements. We recommend prioritizing computational infrastructure by fostering cross-border collaborations, developing shared cloud environments, and establishing standardized procurement frameworks for advanced technology. Such collective efforts could significantly enhance the capacity of statistical agencies across Europe to innovate and adapt in an era of rapidly growing data complexity.

4.2 Data Coverage and Accuracy

- The availability of traffic cameras in Sweden was limited, with most of them located in urban areas, making it difficult to generalize findings.
- At the time the project was running, the Faster RCNN trained model at Data Science Campus had not been tested with adverse weather – such as heavy snow, nor had it been stress tested during hours of darkness. Worth remembering that the daylight hours in Sweden are shorter than the UK during winter and longer during summer. (It is possible that currently, this issue has been resolved at ONS).
- Similarly, it is unknown what impact heavy snow will have on the detection of objects.
- Thus, the machine-learning model used for object detection possibly could face accuracy challenges, particularly in adverse weather conditions and during nighttime.

4.3 Strengths

Despite the challenges faced during implementation, the project showcased several notable strengths. Drawing inspiration from the UK use case, the following key benefits were identified:

- **Efficient Resource Utilization**: The system leverages existing public resources, such as traffic cameras, minimizing the need for additional infrastructure.
- **Cost-Effectiveness**: Cloud infrastructure costs are relatively low compared to traditional methods of traffic data collection.
- **Timeliness**: The system delivers near real-time statistics, with updates available on a daily or weekly basis, ensuring data remains current and actionable.
- **Comprehensiveness**: By using existing public resources, the system can detect a wide range of objects—including cars, buses, and pedestrians—providing a holistic view of traffic patterns.
- **Privacy Assurance**: The system ensures privacy compliance by not storing or analysing personal identifiers, adhering to stringent data protection standards.

4.4 Future Opportunities

To enhance the robustness of the indicators, the following additional data sources were proposed:

- Mobile phone position data for real-time population movement.
- Sensor data, such as congestion tax data and traffic loop data, to complement traffic camera data.
- Smart Devices in Citizen Science - The Telraam Project[2]

The potential of citizen science projects such as the Telraam project has been described before in the ESSnet Big Data II [3]. It exemplifies how smart devices can enhance official statistics through citizen science. It uses low-cost cameras and Raspberry Pi devices with open-source software to collect data on traffic, including pedestrians, cyclists, cars, and lorries. The devices process images locally, calculating counts and speeds before sending anonymized data to a central database. This allows real-time data analysis while ensuring privacy and transparency.

[2] https://telraam.net/

The Telraam approach demonstrates a model of decentralized data processing and citizen engagement. The software's open-source nature fosters trust, as users can verify its function. The concept could be expanded to recognize additional objects or collaborate with other smart devices, making it highly adaptable for statistical observations.

Since its inception, the project has scaled up, covering more cities, and using more advanced equipment. However, the principle of on-device processing remains key, maintaining relevance for future projects like UC6, which could explore and extend this model for further statistical applications.

Another interesting project related to the use-case in this document is the pedestrian count data project as presented by DESTATIS [4]. In this project an R-based processing pipeline has been implemented that on a weekly basis calculates and disseminates the pedestrian count index, working on data from a German startup company 'Hystreet'. This example shows that besides public traffic cameras or citizen science data, another type of data collected by commercial purposes might be of interest.

4.5 Cross-cutting issues

Python Production Environment

Sweden is actively working to establish a Python production environment, as this critical infrastructure is currently lacking. This limitation restricts the scalability and operational efficiency of the system.

Script Longevity

The current Python solution is not suitable for long-running scripts. Relying on a stand-alone laptop to run indefinitely is impractical for a production environment, as it introduces risks such as system interruptions and resource constraints.

Data Loss Risk

There is a notable risk of data loss in the current production setup. For instance, when reviewing images scheduled via Task Scheduler at Webbkameror.se, only an IP address was retrieved. Additionally, since all images from a given camera are accessed through the same URL, it appears that the data provider does not maintain a backup of the images. To mitigate this risk, it is essential to implement regular backups and ensure that images are stored in a persistent and reliable manner.

5 Conclusion

This project set out to explore the applicability of using publicly available traffic camera images to calculate a busyness indicator, a concept successfully piloted in the UK, and to adapt it for use in other countries, with Sweden as a case study. From a technical perspective, the concept proved feasible, as demonstrated by the successful replication of the approach in Sweden. The findings in this document confirm that the methodology is technically portable across countries.

However, practical challenges were identified that could limit the full realization of this concept. The analysis of data collected in Sweden indicates that the system performed reasonably well, as evidenced by the low proportion of non-operational cameras and the predictable patterns observed in the visualized data. These results highlight the potential for leveraging traffic camera data to create busyness indicators in Sweden, mirroring the UK's approach.

Nevertheless, achieving reliable and scalable indicators will require addressing key limitations related to data coverage, accuracy, and accessibility. Sweden's unique conditions, such as its distinct weather patterns, differences in the number and placement of cameras, and evolving societal attitudes toward privacy, have posed challenges. Notably, changes in privacy concerns have reduced the availability of public traffic camera images, underscoring the need for alternative or complementary data sources.

Future efforts should prioritize integrating traffic camera data with additional datasets to develop a more comprehensive and robust picture of urban activity. Potential complementary data sources include:

- **Mobile phone positioning data** to capture population movement patterns.
- **Sensor data**, such as congestion tax information or traffic loop measurements, to provide context for vehicular activity.
- **Citizen science initiatives**, such as smart device data collection, exemplified by the Telraam project discussed in Section 4.4.

Joining multiple data streams, it is possible to overcome current limitations and enhance the reliability and utility of busyness indicators. These advancements will better support decision-making processes and provide valuable insights into urban dynamics, not just in Sweden but in other regions seeking to adopt similar methodologies.

6 References

[1] Chen, L.; Grimstead, I.; Bell, D.; Karanka, J.; Dimond, L.; James, P.; Smith, L.; Edwardes, A. Estimating Vehicle and Pedestrian Activity from Town and City Traffic Cameras. Sensors 2021, 21, 4564. https://doi.org/10.3390/s21134564

[2] https://datasciencecampus.ons.gov.uk/projects/estimating-vehicle-and-pedestrian-activity-from-town-and-city-traffic-cameras/

[3] ten Bosch, O., Quaresma, S., De Cubellis, M., „Deliverable L3: Description oft he findings regarding Task 3, Smart Devices", WP L (Preparing Smart Statistics), ESSnet Big Data II, 30 Oct 2019

[4] Buhl Y.: New Data Sources and Off cial Statistics, an Exemplary Research Workflow in R Using Different API Wrapper Packages. Presentation at the uRos 2024 Conference, November 2024, Athens