Case study: Border survey for foreign travellers in Finland

Ella Oksa, senior statistician, Statistics Finland, ella.oksa@stat.fi

Ossi Nurmi, senior statistician, Statistics Finland, ossi.nurmi@stat.fi

Abstract

This paper provides a case study on a new border survey for foreign travellers in Finland. Visit Finland, the national tourism promoting organization, launched the survey in March 2023 to collect data on the trips, expenditure, and characteristics of foreign visitors in Finland. The objective of the border survey is to provide reliable and comprehensive information about the volume, economic significance, regional structure and characteristics of inbound tourism in Finland to support the industry and political decision-making. The survey provides basic information on the demographic characteristics of foreign travellers, their expenditures related to the trip, timing and duration of trips, destinations within Finland, purpose of the trip, types of accommodation and other travel characteristics. The survey uses various selection criteria and methods to form the sample from the population that is travellers who leave Finland by flight or ferry. The survey collects data from passengers through interviews conducted prior to departure. The results of the survey are published through Visit Finland's Rudolf database. The processed survey data set is available as open data. Additionally Visit Finland publishes a visual and interactive dashboard based on the open data. We present the methodological choices and feasibility of the chosen methods for different conditions. Furthermore, we discuss potential limitations as well as additional data sources for improving the survey results.

Introduction

Tourism is a significant sector of the economy in Finland, generating about 2.7% of the gross domestic product in 2019 and employing about 5.8% of the workforce before the pandemic (Finland's tourism accounts for 2020-2021, 2023). Understanding the economic impacts of tourism is essential for supporting the industry and political decision-making.

Visit Finland, the national tourism promoting organization, launched a new border survey in March 2023 to collect data on the visits, expenditure, and characteristics of foreign visitors in Finland. The data is obtained by interviewing travellers at the point of departure from Finland, such as airports and ports. The objective of the border survey is to provide reliable and comprehensive information about the volume, economic significance, regional structure and characteristics of inbound tourism in Finland to support the industry and political decision-making. The information is needed to allow for the better targeting of tourism marketing, for the development of tourism services and the preparation and evaluation of strategies. The information is utilized by regional tourism organizations, tourism companies, Visit Finland, Statistics Finland, municipalities and ministries. (Koskela, et al., 2021)

Visit Finland collaborates with Statistics Finland, Norstat Finland Oy and Visitory Oy (formerly TAK Oy). Statistics Finland plans the sample, processes the data for analysis, and publishes the data as database tables and open data. Norstat handles the field work: recruiting, training and supervising interviewers, and controlling the quality of interviews. Visitory takes care of the planning and technical requirements for data collection, such as translating and implementing the questionnaires. (Visit Finland, 2024)

The results of the survey are published monthly in Visit Finland's Rudolf statistical database and a short report is published in Visit Finland's website in each month. The survey data is also published monthly as open data in Finnish open data portal avoindata.fi.

Methods

The border survey is a systematic and comprehensive study of the travel and spending patterns of foreign travellers in Finland. The survey uses various selection criteria and methods to form the sample from the population that is travellers who leave Finland by flight or ferry. The data collection method of the survey is a personal interview, in which the interviewer records the target person's answers on a data collection form using a tablet. (Visit Finland, 2024), (Visit Finland; Statistics Finland, 2022)

Target population and sampling

The aim of the survey sampling is to form a representative and random sample of the population. Thus, the sampling aims to give as unbiased a picture as possible of the distribution of trips in the target group according to the traveller's country of residence, purpose of the trip, time of travel and other trip characteristics.

The population from which the sample is drawn consists of travellers departing Finland via various border crossing points. The target population, which is the focus of the study, is the subset of travellers who are permanent residents of foreign countries, thus excluding those who are permanent residents of Finland, from the target population.

Selection of border crossing points

In the sampling design phase conducted during 2022, the main task was to select the border crossing points to be included in the survey. The passenger statistics for the border crossing points are compiled Statistics Finland, Finnish Border Guard and Finnish Customs compiles passenger statistics. However, the main difficulty of these statistics is that, for the most part, they do not differentiate the passengers to Finnish residents and residents of other countries. We estimated the percentage of non-resident travellers based on seasonal variations, flight destinations and types (scheduled or charter) as indicators.

Land crossing sites also provide challenges, especially those located on the Schengen borders with Sweden and Norway. The vehicles do not stop at these borders and conducting interviews there is either very difficult or not possible at all. The eastern border with Russia was no longer relevant for tourism after Russia invaded Ukraine in February 2022 and as a consequence Finland ceased to provide tourist visas to Russian citizens.

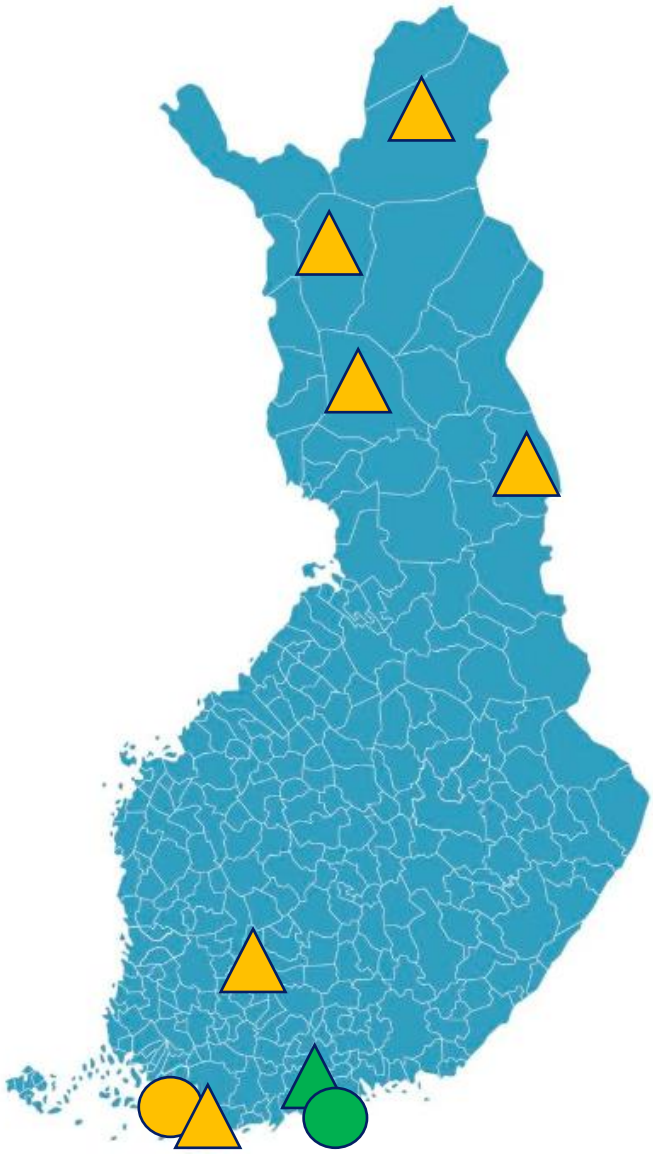

Figure 1. Selected border crossing points.

The selected border crossing points consist of seven airports (triangles) and two ports (circles). These cover approximately 94% of all departing air and ferry passengers based on 2022 passenger statistics. Interviews are conducted year-round (green color) at the Helsinki airport and ferry terminals and during the main season at the remaining sites.

Sampling plan

At the ports, the interviews are conducted both inside the passenger terminal as well as the car lines. At the airport, the passengers are interviewed at the departure gates while they are waiting for boarding the aircraft. The objective of the sampling is that each passenger would have roughly an equal probability of being selected for an interview. This requires systematic planning on a quarterly and monthly level as well as in the day-to-day fieldwork carried out by the interviewers.

Statistics Finland is responsible for providing the monthly sampling plan. For airports, the plan is based on the number of seats available. This data is provided monthly by Finavia, a public limited company responsible for Finland's airport network. Domestic flights are excluded as well as destinations such as Canary Islands or Greek Islands, that are primarily holiday destinations for Finnish residents.

There are two main dimensions or quotas in the plan: time and flight destinations. For the time dimension, the number working shifts available are allocated to weekdays and time slots within in the day based on the number of departing seats. In the following example for Helsinki airport, there are 43 shifts available for each month in the 3rd quarter of 2024. First, the shifts are allocated to weekdays and within each weekday, to morning-, day- and evening shifts. Norstat, the company responsible for the interviews, will allocate plan the actual working shifts according to this plan.

Table 1. Allocation of working shifts for Helsinki Airport.

Working shifts	Month			
	Jul-2024	Aug-2024	Sep-2024	Total
Monday	**7**	**5**	**9**	**21**
1. Morning (6-12)	3	2	3	8
2. Day (12-18)	3	2	4	9
3. Evening (18-01)	1	1	2	4
Tuesday	**7**	**6**	**8**	**21**
1. Morning (6-12)	3	2	3	8
2. Day (12-18)	3	3	4	10
3. Evening (18-01)	1	1	1	3
Wednesday	**7**	**5**	**5**	**17**
1. Morning (6-12)	3	2	2	7
2. Day (12-18)	3	2	2	7
3. Evening (18-01)	1	1	1	3
Thursday	**6**	**7**	**4**	**17**
1. Morning (6-12)	2	3	1	6
2. Day (12-18)	3	3	2	8
3. Evening (18-01)	1	1	1	3
Friday	**6**	**7**	**4**	**17**
1. Morning (6-12)	2	3	1	6
2. Day (12-18)	3	3	2	8
3. Evening (18-01)	1	1	1	3
Saturday	**5**	**7**	**4**	**16**
1. Morning (6-12)	2	3	1	6
2. Day (12-18)	2	3	2	7
3. Evening (18-01)	1	1	1	3
Sunday	**5**	**6**	**9**	**20**
1. Morning (6-12)	2	2	3	7
2. Day (12-18)	2	3	4	9
3. Evening (18-01)	1	1	2	4
Total	**43**	**43**	**43**	**129**

For the flight destinations, a quarterly quota is established where each flight destination gets a target share as a percentage of all interviews during that quarter based on available seat capacity. For example, in case 5 percent of all seats are departing from Helsinki airport to Stockholm Arlanda airport the target is to collect 5 percent of all interviews from flights bound to Stockholm Arlanda in that quarter. Norstat

monitors the accumulation of interviews by destination as the quarter progresses and targets those flights that are falling behind of the quarterly quota.

For ferries departing to Estonia, Sweden and Germany, a similar approach is used but instead of seats, the plan is based on the estimated number of passengers for each ferry departure. For this purpose, we use actual passenger data from Fintraffic, a state-owned traffic control and -information company, and combine this with the ferry schedules. The data from Fintraffic provides the number of passengers for each ferry departure and this can be used as a proxy for ferry departures in the future. The number of passengers have a high monthly seasonality, with July and August being the peak months. There is also a weekly repeating cycle, where ferries have a significantly higher passenger volume on Friday, Saturday and Sunday departures and less passengers from Monday through Thursday. Early morning and late evening ferries also tend to have a lower passenger volume than other departures. The sampling considers the monthly seasonality, weekly cycle and different times of day, and provides a monthly plan similarly based on the estimated number of passengers.

Selection of persons for interview

During the working shift, the interviewers pick respondents according to predetermined sampling intervals to ensure random selection. The intervals, for example every 10th person, are assigned based on the size of the aircraft or ferry. The ideal sampling interval should be low enough, so that interviewer will not run out of potential respondents, but also high enough, so that each passenger has a chance of being interviewed. Typically, the sampling interval ranges between 3 (small aircrafts) and 10 (ferries and large aircrafts).

An additional sampling layer is needed for motor vehicles queuing to board the ferry. For cars (and other vehicles), the interviewer randomly selects a passenger inside the vehicle for an interview. A practical solution for this randomization is to take the last number in the vehicle's license plate (0 being ten) and pick the n:th passenger inside the vehicle for an interview.

The questionnaire

The survey provides basic information on the demographic characteristics of foreign travellers, their expenditures related to the trip, timing and duration of trips, destinations within Finland, purpose of the trip, types of accommodation and other travel characteristics.

At the outset of the questionnaire, a series of screening questions are posed to ascertain the respondent's eligibility for the target group. The screening questions are related to the flight destination, country of residence and age. If these initial inquiries reveal that the individual does not belong to the target group, the interview is terminated. To address the confusion experienced by both interviewers and respondents due to the abrupt end of the interview, additional questions were introduced for Finnish residents. This measure was taken to facilitate a more conventional closure of the interview. The added questions aimed to obtain information regarding the primary purpose of the trip and destination country.

If eligibility is confirmed, the questionnaire offers three options for the primary purpose of the trip: transit, personal, and work-related. For transit passengers, the questionnaire is brief, consisting only of a question about expenditures, after which it concludes. For personal and work-related trips, the questionnaire continues.

The questionnaire is divided to the following themes:

- Subject background information: age, gender, travelling group, nationality, country of residence and for some countries of origin, also area of residence.
- Main purpose of the trip: personal, work-related, transit)

- More detailed purpose of personal and work-related trips, for example recreation, meeting friends or relatives, studying, participating in a conference or congress
- Date and duration of trip: start date, end date (i.e., interview date), duration of trip in nights
- Mode of transport: outward and return journey (ferry or flight), over 50 km journeys within Finland
- Travel destinations: main destination and other destinations in Finland, other destination countries in addition to Finland
- Accommodation and booking: main type of accommodation, package trip, month of booking.
- Expenditures: cost of package (in case of package trip), advance payments before arriving in Finland and costs paid in Finland which include accommodation, restaurants, fuel, getting around in Finland, culture and sports, other expenditure including shopping
- The share of expenditure paid in cash
- Activities or experiences in Finland, the respondent is asked to select up to three options from the following list:
    - Nature experiences
    - Physical or sport activities outdoors
    - Wellbeing and relaxation
    - Cultural experiences
    - City break
    - Participating in a cultural or sport event
    - Shopping
    - Touring, road trip
    - None of the above

In addition, some technical information is collected for the data processing phase, such as the time and place of the interview and the destination of the flight or ferry (Visit Finland; Statistics Finland, 2022). The average duration of the interview for foreign respondents is approximately 8 minutes and 20 seconds.

Data processing and quality

Visitory, the company that implements the data collection technically, submits the raw interview data to Statistics Finland every month. Statistics Finland processes the raw data into a representative dataset that describes the entire target population and is ready for analysis. The data process involves adding weight to each respondent to indicate how many passengers in the target population they represent. The data is classified with the necessary categories such as the municipal and provincial classifications of destinations based on administrative regions and grouping the detailed information into broader categories such as the respondent's age into age groups. Derived information is added from the responses to the data such as the carbon footprint of the trip. The missing data such as partially missing data on expenditures is completed with statistical methods for example by filling in the missing data with data from other similar trips. Missing data are common in all surveys and completing them is essential for the data's representativeness.

Editing

The date of the interview is automatically recorded, and the respondent provides the date of their arrival in Finland. Additionally, the respondent reports the main destination and up to three other overnight destinations they visited in Finland and the number of nights spent at each destination. Ideally, the difference in nights between the date of arrival and the interview date should equal the sum of the nights spent at the reported destinations.

Respondents can choose destinations from a predefined list or provide open-text responses. The list includes municipalities in Finland, recreational locations such as golf courses, ski resorts and national parks. However, open-text responses often require significant editing due to misspellings or the inclusion of

places not located in Finland. While most places are recognized programmatically, some manual editing is necessary. This stage also involves identifying transit passengers who may have incorrectly reported "transit" or "transfer" as a destination instead of selecting "transit" as their primary purpose. It is important to accurately identify transit passengers, as they are excluded from the final reported figures and the published open dataset. Similarly, crew members are excluded from the final data.

The process identifies destinations outside Finland, and the nights spent outside Finland are excluded from the calculation of the length of stay. The length of stay calculated from the difference of arrival and departure dates is compared with the sum of nights spent at each destination. If there is a discrepancy, the data is matched by adjusting the number of nights reported, either by adding or subtracting the appropriate number of nights based on the distribution of reported nights.

Respondents may report only their own expenses or the expenses of several persons, for example the whole family or travel group. Respondents can report their expenses in a currency of their choice. During the data process, monthly exchange rates against the euro are retrieved from the European Central Bank API to convert reported expenses into per-person expenses in euros. Furthermore, the total costs per day are calculated based on the length of stay. At this point, observations with expenses exceeding plausibility thresholds are set to missing and will undergo the imputation process. Plausibility thresholds are set respectively for each expenditure category.

Imputing missing expenditure and unbundling of package trips

The questionnaire has only a few required fields, these include respondents age and country of residence, and destination of flight or ferry. Questions about expenditures, destinations in Finland, and other trip characteristics are optional and may be left unanswered for various reasons, such as the respondent's lack of knowledge, ability, or willingness to answer.

For the expenditures, missing data at the observation level are imputed by using the mean expenditure of a suitable reference group. The reference groups are formed based on the analysis of the data and the background factors that affect the expenditures. For instance, the daily expenses on accommodation and restaurants tend to be higher for work-related trips than personal trips.

Package trips and advance payments usually include various products or services that need to be unbundled to expenditure categories because the price of flight or ferry tickets is out of scope and needs to be excluded. For package trips, respondents report the services included in the package, which always assume flight or ferry tickets and accommodation. Additional services may include meals, entrance tickets, other services in Finland, and services in other countries. This information is used to unbundle the package trip expenditure.

The costs for accommodation, meals, culture and sports, and other categories are imputed for package trips, using average costs from similar types of trips. These imputed costs are then compared with the price of the package. If there is a significant difference, the costs are adjusted by either subtracting or adding an appropriate amount to align with the package trip expenses.

Carbon footprint

In the survey, the essential information for calculating the carbon footprint includes the country and region of residence, the duration of the trip and the mode of transport for the round trip. Furthermore, the carbon footprint is impacted by the type of accommodation and the modes of transport used to reach the primary destination and other destinations in Finland. For trips with private cars, also the number of travel companions impact the CO2 emissions per person. If multiple destinations are visited in Finland, the total journey length within Finland is calculated by averaging the minimum and maximum lengths of all possible routes. (Nurmi, 2023)

The calculation model is conceptually based on models used by Switzerland Tourism and Innovation Norway (Grythe & Lopez-Aparicio, 2021). These models combine the characteristics of the trip with various emission factors, such as CO2 per passenger kilometer and CO2 per hotel night.

The accuracy of the calculated results is predominantly influenced by the emission factors employed. Emission factors also change over time and need to be updated regularly. For instance, modern aircraft are more efficient compared to their predecessors, a trend also in sea transport. Similarly, the per-kilometer emission rates for cars are on a decline, with electric vehicles producing no direct emissions. The indirect emissions of electric vehicles depend on the methods employed in electricity production (Nurmi, 2023). The table below presents the emission factors currently in application.

Table 2. The emission factors used to calculate the carbon footprint of a trip.

Topic	Subject	Unit	Coefficient	Source
Air Travel	Flight < 200 km	g CO2 / passenger-km	167	Finnair
Air Travel	Flight 200 - 500 km	g CO2 / passenger-km	133	Finnair
Air Travel	Flight 500 - 800 km	g CO2 / passenger-km	108	Finnair
Air Travel	Flight 800 - 1 000 km	g CO2 / passenger-km	87	Finnair
Air Travel	Flight 1 000 - 5 000 km	g CO2 / passenger-km	78	Finnair
Air Travel	Flight > 5 000 km	g CO2 / passenger-km	54	Finnair
Ferry Travel	Helsinki - Tallinn	g CO2 / passenger-km	74	Finnish Shipping Companies
Ferry Travel	Helsinki - Stockholm, Turku-Stockholm	g CO2 / passenger-km	54	Finnish Shipping Companies
Train	Long-distance Train	g CO2 / passenger-km	1.5	LIPASTO 2019
Bus	Bus	g CO2 / passenger-km	53	LIPASTO 2019
Car	Car (Diesel or Gasoline)	g CO2 / driving-km	150	LIPASTO 2019
Accommodation	Hotel Night	g CO2 / room-night	14000	DEFRA (Europe)
Accommodation	Other than Hotel Accommodation	g CO2 / room-night	10000	Assumption
Dining	Meal	g CO2 / meal	2000	Switzerland Tourism

Example of a trip with a carbon footprint of 850 kg CO2
A round trip from California, United States, to Helsinki, Finland with a stopover in New York, United States results in a carbon footprint of 491 kg CO2. The travel chain within Finland, from Helsinki to Inari and returning to Helsinki, results in a carbon footprint of 212 kg CO2. Staying for 7 nights in hotels in Finland, specifically in Inari and Helsinki, and including meals, results in a carbon footprint of 146 kg CO2.

The following images illustrate the computations involved in the travel chain from California, United States, to Helsinki, Finland, as well as the internal travel within Finland. This travel chain covers the round trip from the country of origin to the destination and the return journey.

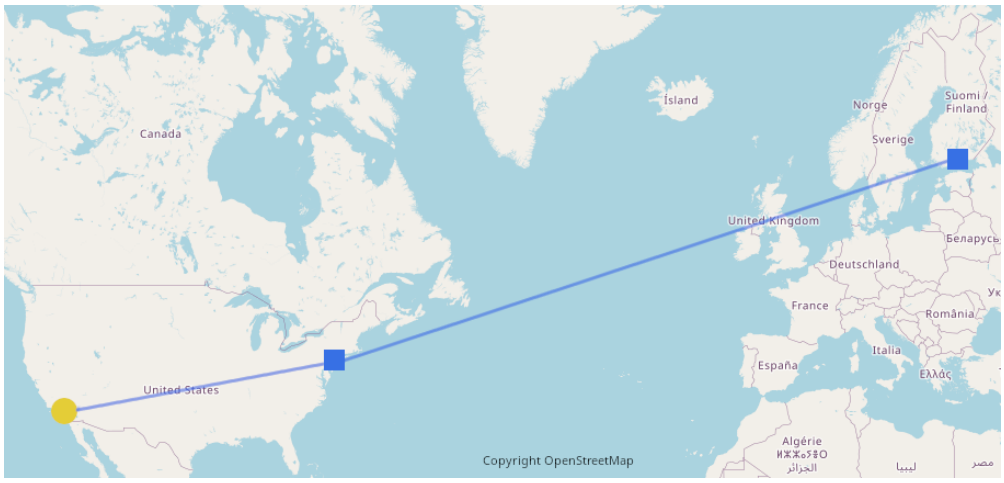

Figure 2. Flight from United States to Finland with a stopover in New York. Map data Copyright OpenStreetMap contributors.

Figure 3. Flight from Helsinki to Inari. Map data Copyright OpenStreetMap contributors.

Weights

Survey weights are used to generalize the data from a single respondent to the entire population. The weighting coefficients are calculated by the destination of the flight or ferry and by the time interval, either month or quarter. The computation involves the sample size (n) and the population size (N). The sample includes the people interviewed in the interval, and the population includes the total number of passengers

in the same interval. The data concerning the aggregate number of passengers, segmented by flight or ferry destination, is sourced from Air Transport Statistics and Foreign Shipping Traffic Statistics, respectively.

The following hypothetical example illustrates the calculation principle of the weighting coefficient:

> Suppose that the total number of departing air passengers from Helsinki to Stockholm in March was 100,000 (N). In March, 100 (n) interviews were conducted at Helsinki Airport with travellers on a flight to Stockholm. Among the respondents, 50 were residents of Finland, who were part of the population of the study but not the target population. The other 50 were not residents of Finland and therefore part of the target population. All these respondents are assigned a weighting factor of N/n=1,000 in the data.

A similar weighting is applied separately to all flight and ferry destinations. The destination of the flight or ferry is only used for the weighting and is removed from the data before analysis. In the data, visitors are categorized by their country of residence, and it is not possible to obtain the number of passengers by destination.

Determining the ferry destination is straightforward, as it can be deduced from the port. The initial assumption of the flight destination is based on the flight number shown at the departure gate. The interviewer confirms the flight with the respondent and records the flight number and destination. If inconsistencies arise, editing is done programmatically to match flight numbers and destinations, assuming most respondents fill in both fields correctly. Manual editing is only required if no flight number is provided, and the destination cannot be verified through airport abbreviations.

During the weight calculation phase, it is verified that all observations have a weight coefficient. If any observations lack a weight coefficient, it indicates that there was no corresponding flight to the claimed destination in the departure data. These observations are manually checked to find the correct destination.

Calculating the weights

Weights are calculated on a monthly and quarterly basis. The surveys sampling design for Helsinki airport is designed to ensure that, ideally, interviews are collected in proportionate numbers from all flight destinations included in the survey over the course of a quarter. Consequently, there may be instances within a single month where certain flight destinations, although part of the sample, do not have any interviews collected. In such cases, the passenger numbers for these destinations are redistributed to those destinations from which interviews have been gathered, in accordance with their respective shares. Therefore, the additional number of passengers allocated to each destination with interviews depends on the volume of passengers for that particular destination. For destinations with a larger passenger count, a correspondingly greater number of passengers is distributed.

To give detailed explanation of the calculation of the weights in one airport, let's assume we have k flight destinations with interviews. Let n_i be the number interviews conducted per destination, and N_i be the number passengers per destination, index i signifying the flight destination, $i = 1, 2, \dots, k$. Let N be the total departing passenger count and S equal the sum of passengers to the destinations with interviews, that is $S = \sum_{i=1}^{k} N_i$.

Now the passenger count to allocate is $N - S$. Allocated passenger count per destination, a_i, can be calculated as follows: $a_i = (N - S) \times (N_i/S)$. To obtain the adjusted passenger count, the allocated passenger count is added to the original passenger count. With this information the weight per destination can be calculated by dividing the adjusted passenger count of corresponding destination with the number of interviews conducted among passengers to that destination, that is $w_i = (a_i + N_i)/n_i$.

To illustrate, consider a simplified hypothetical example of passenger distribution at Helsinki airport. Imagine that in one month, Helsinki airport has a total of 100 000 departing passengers. During this month, 100 interviews are conducted at Helsinki airport. The distribution of departing passengers and the interviews collected is as follows:

Table 3. The simplified hypothetical example of passenger and interview distribution of one month in Helsinki Airport.

Airport Name	Passengers	Interviews
Stockholm Arlanda Airport	40 000	40
London Heathrow Airport	30 000	40
Charles de Gaulle Airport	20 000	20
Amsterdam Schiphol Airport	5 000	0
Frankfurt am Main Airport	5 000	0

If weights for each destination were calculated by simply dividing the total number of passengers by the number of interviews (e.g., Stockholm Arlanda Airport: weight = 40 000 / 40), this method would overlook the passengers from destinations without interviews (Amsterdam Schiphol Airport, Frankfurt am Main Airport). To address this, passengers from destinations lacking interviews will be distributed among those with interviews. This ensures that the interviewed respondents also represent these passengers. The adjusted passenger count for each destination accounts for the original number of passengers plus an additional allocation proportional to their relative size.

Now in our example we have 90 000 passengers to destinations with interviews and 10 000 passengers to destinations without interviews. We adjust the passenger distribution as follows:

- 10 000 * (40 000 / 90 000) = 4 444 passengers are allocated to Stockholm Arlanda Airport, resulting in an adjusted passenger count of 44 444 (4 444 + 40 000)
- 10 000 * (30 000 / 90 000) = 10 000 passengers are allocated to London Heathrow Airport, resulting in an adjusted passenger count of 33 333 (3 333 + 30 000)
- 10 000 * (20 000 / 90 000) = 2 222 passengers are allocated to Charles de Gaulle Airport, resulting in an adjusted passenger count of 22 222 (2 222+ 20 000)

Subsequently, the weight per destination is calculated using these adjusted figures:

- Stockholm Arlanda Airport: weight = 44 444 / 40
- London Heathrow Airport: weight = 33 333 / 40
- Charles de Gaulle Airport: weight = 22 222 / 20

We see that the sum of the adjusted passenger counts is equivalent to the total number of passengers, hence all passengers within the sample have been considered.

For smaller airports in the northern regions namely Ivalo, Kittilä, Kuusamo, and Rovaniemi, the weight calculation is adapted to use the country of the flight's destination rather than the specific airport. This adjustment is due to the high volume of charter flights during the winter season in Lapland, which results in a distribution of passenger counts across smaller airports. Interviews are collected from a random selection of destinations, leading to potentially high allocation numbers. Therefore, considering the dispersion of passengers across numerous small destinations, it is more practical to base the calculation of weights on the destination country.

For ferry destinations, the weight calculation is more straightforward. Given the limited number of possible destinations and the collection of interviews from each destination every month, there is no need to allocate passenger counts.

Sample size

The data collected from March 2023 to April 2024, shows that the share of the target group, non-resident visitors to Finland, varies between months. In March 2023, the non-target group constituted 61% of the interviews, while the target group represented 39%. This pattern, where the non-target group maintains a larger share, persists across most months, with the target group's share typically ranging from 25% to 40%.

Notably, during the peak summer month of August 2023, and the winter months of December 2023 through March 2024, the target group's share surpassed 50%. This type of seasonality is expected as it reflects the holiday seasons.

Throughout the period from March 2023 to April 2024, the average number of valid interviews was 1513 interviews per month. The minimum recorded in any single month was 836 interviews in March 2023, and the maximum was 2262 interviews in July 2023. The target group's lowest interview count was 324 in April 2024, while its peak was 1069 in December 2023. During the 12 month-period from March 2023 to February 2024 a total of 7582 target group interviews were conducted.

In the survey data spanning from March 2023 to April 2024, the lowest monthly margin of error in the total number of trips was in December 2023 at 7%, while the highest margin of error was in April 2024, reaching 15%. The monthly margin of error is directly related to the sample size and typically sets around 10% of the total number of trips. For the quarterly analysis, spanning from the second quarter of 2023 to the first quarter of 2024, the margin of error for the total number of trips fluctuated between 4.5% and 5.3%.

Results

The results of the survey are published through Visit Finland's Rudolf statistical database, where two tables are updated every month and seven tables every quarter. The processed survey data set is available as open data in open data portal avoindata.fi. Additionally, on the home page of the Border survey (Visit Finland, 2024), Visit Finland publishes a visual and interactive dashboard based on the open data.

The monthly results available in Rudolf database provide information on key figures of foreign visits to Finland, such as purpose, means of transport, stay duration, visitor numbers, expenditure, and trip-related carbon emissions. Additionally, a table is updated monthly with a rolling dataset of the past 12 months. The sample size of this table enables detailed reporting on the visitors' countries of origin and their home regions. As previously mentioned, both transit passengers and crew members are omitted from the reported statistics. Consequently, the term 'foreign visitor' remains accurate. In the expenditure reported the price of flight or ferry tickets is excluded.

The quarterly results available in Rudolf database provide more detailed information on the characteristics and behavior of foreign visitors to Finland. They include finer classifications by the visitor's country of residence and the main destination in Finland. They also include more information on the expenditure categories and the other destinations besides the main one in Finland.

The dissemination of results through the Rudolf database is intended for those seeking to follow the figures via a web user interface. In contrast, the availability of open data serves those aiming to utilize the data for example creating dashboards, compiling reports, or conducting research. Potential users of this data include Visit Finland, Statistics Finland, destination management organizations, academic institutions, researchers, and companies. For example, Statistics Finland will utilize data from border survey to determine the inflows (credits) in the calculation of the travel item in Balance of payments statistics.

Visit Finland has released an extensive dashboard that presents a diverse range of metrics, offering insights from destinations, travellers profile and tourism flows. The dashboard also includes information of the

departure countries of visitors and their chosen destinations. The key statistics are displayed on the report's homepage.

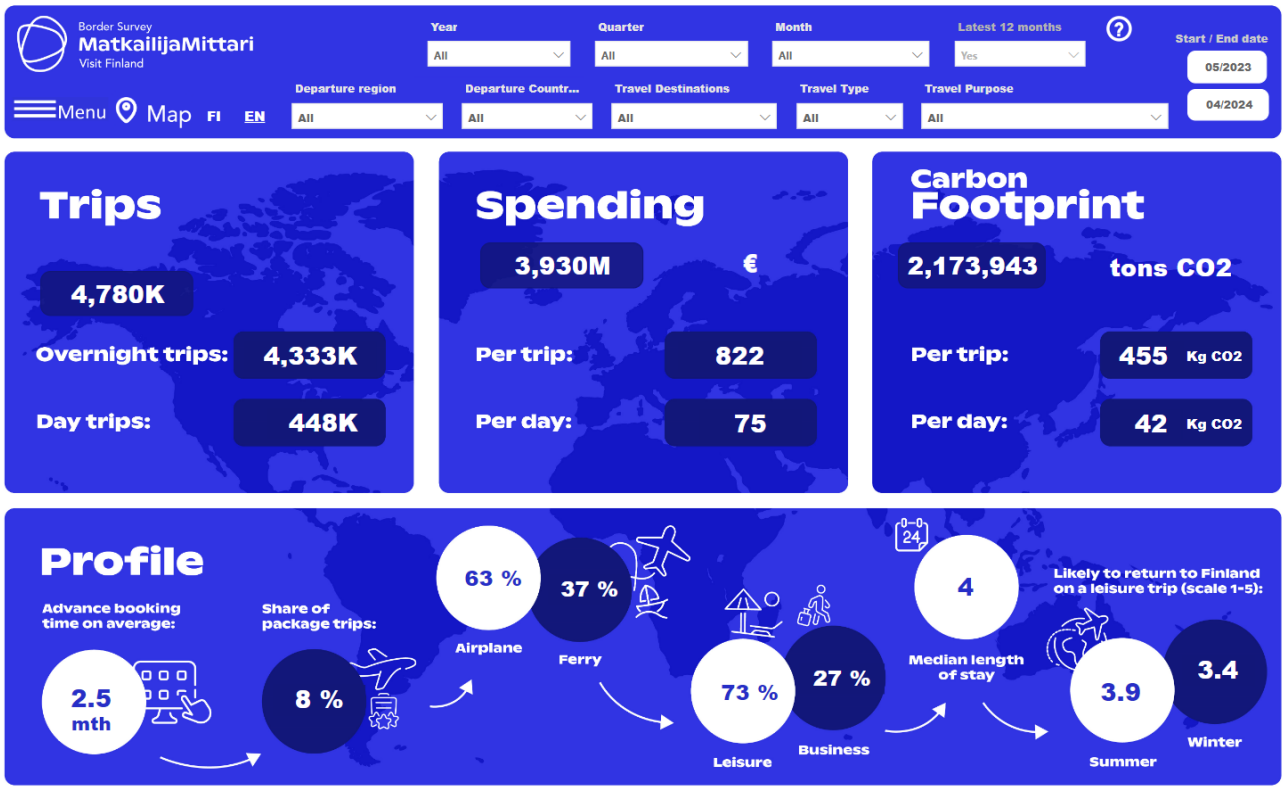

Figure 4. Front page of the interactive report released by Visit Finland, which utilizes open survey data as the primary source.

Between May 2023 and April 2024 foreign travellers made 4.8 million trips to Finland and spent the total of 3.9 billion euros. Average expenditure per trip was 822 euros. Average carbon footprint per trip was 455 kg of CO2, consisting of the transport to and from Finland, transport within Finland, accommodation, and meals.

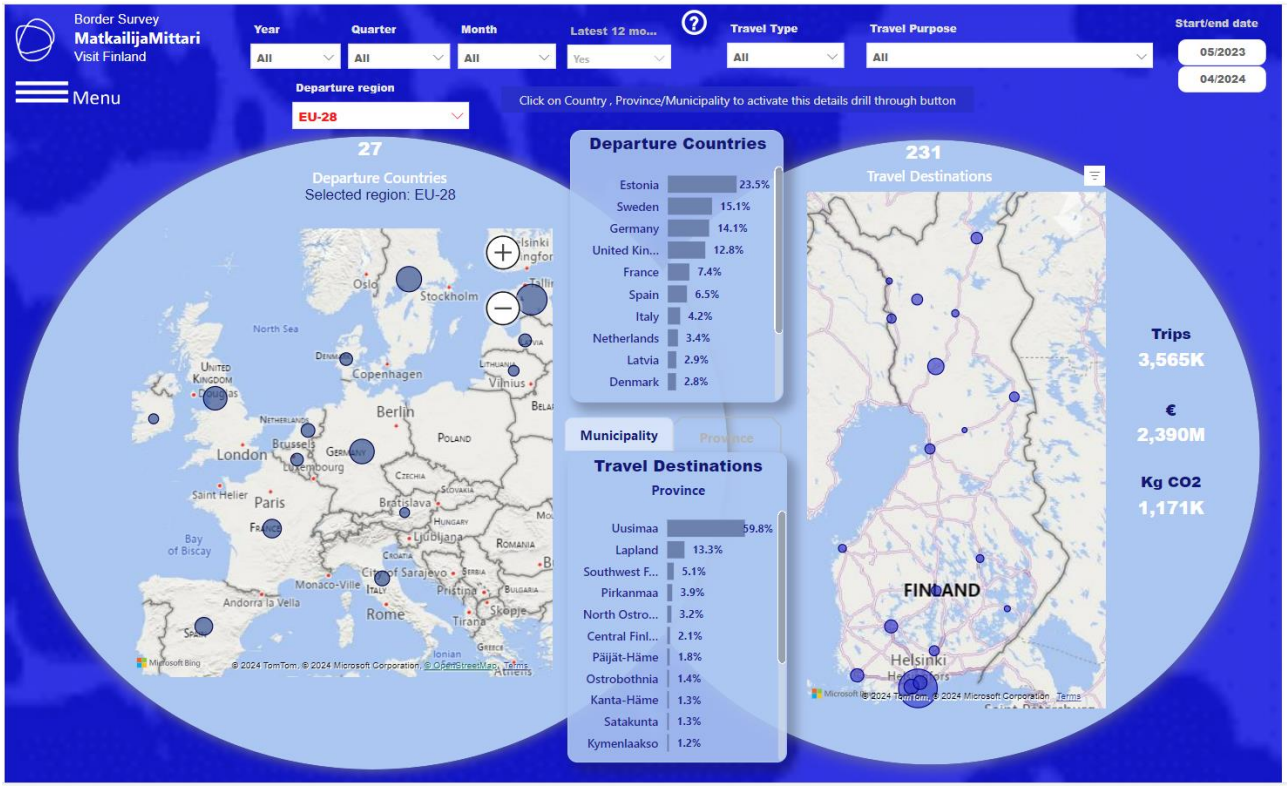

Figure 5. Page of the interactive report released by Visit Finland, presenting visitor origin countries and primary destinations in Finland.

During the 12 months period from May 2023 to April 2024, Estonia, Sweden, and Germany were the primary countries of origin for foreign visitors to Finland, complemented by significant number of trips from the United Kingdom, United States, and France. These countries represented 65% of all international visits. Uusimaa, home to the capital, Helsinki, was the top destination, with a 60% share of all travellers. Lapland followed second, being the main destination for 13% of travellers.

Discussion

Border surveys are common in many countries to obtain information about the economic significance of tourism and the travel behavior of visitors. Border surveys usually have similar principles, such as complementing the interview data with the data on the number of passengers, to generalize the survey data to the population level. The method of interviewing passengers departing the country is suitable for countries with limited number of relevant border crossing points or passport controls, as this allows for random sampling and aims for high response rates.

Challenges and sources of uncertainty

The border survey has some recognized challenges, limitations and sources of uncertainty. Pace of collecting the data is rather slow, as the interviewers can only conduct a limited number of interviews per day due to the challenging locations and very limited time windows for the interviews. The main challenges and sources of uncertainty can be broadly broken down to these categories

- Sample size and granularity of results
- Sampling biases resulting in over- or underestimation
- Missing data
- Quality of responses

Sample size is mainly determined by the survey budget and data collection method. The current sample size, consisting of roughly 8 000 target group interviews per year, is sufficient for the national level, main destination cities and regions and main countries of origin. More granular aggregates are subject to high error margins and uncertainty as is the case with all surveys. Although there is a clear need for more regional data, the marginal cost of increased granularity is currently not feasible or even justifiable. Using additional data sources, such as payment card data or mobile positioning data, is most likely the key for improving granularity.

Potential sampling biases, leading to over- or underestimation, can take place at any stage of the sampling process. There is already an inherent structural bias in the selection of border crossing sites due to the exclusion of road border crossings. This results in underestimation of visitors from and touring trips to neighboring countries Sweden and Norway. Other sources of sampling bias can be related to the allocation of working shifts, sampling of flight destinations, (non-random) selection of passengers for interview. In general, the biases are minimized by commitment to the sampling methodology and rigorous execution of the sampling plan. Some types of biases are, by nature, beyond the control of the survey staff. For example, the passengers who arrive at the gate very late, have a smaller chance of being interviewed. This may be the case for passengers staying at airport lounges or cafes until boarding begins.

Missing data is an issue in all surveys and the border survey is no exception. The missing values need to be imputed and the choice and fine-tuning of the imputing method is crucial for the quality of the final data. For this border survey we have employed a nearest neighbor method for imputing missing expenditure. In this method, each trip is stratified according to trip characteristics that typically have an impact spending, such as trip duration, purpose and type of accommodation. The nearest neighbor method is straightforward and provides adequate results. There are also other imputing methods, for example multivariate imputation and machine learning -based methods, which could be explored to further improve the results.

In contrast to missing data, quality of the responses is related to the interpretation of the questionnaire by the respondents who are from various language and cultural groups. Although the questionnaire has been translated to close to 15 languages, it is very difficult to ensure that the questions are understood correctly and consistently by all respondents. It's also possible that the respondents do not know or remember the answers to all questions but may still provide a response. There are two main ways to deal with these issues. Firstly, the questionnaire should be systematically tested with real respondents. This often reveals which questions are most difficult to interpret and they can be improved before the actual data collection. Secondly, it's important to introduce plausibility checks to detect outliers and other quality issues in the collected data. Once these are found, it's possible to correct these either manually or, in best case, using imputation methods to provide more plausible values.

Additional data sources

The Ministry of Economic Affairs and Employment of Finland's research design report (Koskela, et al., 2021) mentions the potential integration of supplementary data sources. Such data may include transaction records from the Bank of Finland or various credit card organizations, and data from mobile operators. Additionally, commercial datasets from flight booking systems could be considered.

However, the application of these supplementary data sources is not without complexity due to significant conceptual differences. For instance, transaction data may suffer from undercoverage, as it does not account for cash transactions and may omit a share of prepayments. Similarly, data from credit card companies may not be representative of the entire population, but only of their clientele. Overcoverage is another concern, encompassing transaction categories outside the survey's scope and non-tourism-related

expenditures for example by foreign students or immigrants. Mobile operator data is also limited in representation, reflecting the customer base of a particular provider rather than the entire population.

Despite these challenges, the value of additional data sources is recognized. In some nations, border survey data is completed with alternative datasets to derive more accurate estimates. For instance, Statistics Canada has implemented a Small Area Estimation model to complement the travel survey, generating reliable estimates for smaller regions where the survey sample size is insufficient (Statistics Canada, n.d.). Further research is required to integrate additional data into Finland's border survey. However, there is a genuine need for regional results and the additional data sources could provide a more insight to these questions.

References

Grythe, H., & Lopez-Aparicio, S. (2021). The who, why and where of Norway's CO2 emissions from tourist travel. *Environmental Advances*. doi:https://doi.org/10.1016/j.envadv.2021.100104

Koskela, K., Harvio, S., Rantala, T., Isola, R., Vujicic, S., & Santasalo, T. (2021). *Report on the research design examining the spending and travel behaviour of foreign tourists.* Ministry of Economic Affairs and Employment of Finland. Retrieved January 2024, from http://urn.fi/URN:ISBN:978-952-327-709-0

Nurmi, O. (2023). *Hiilijalanjäljen laskenta Matkailijamittarissa.* Visit Finland. Retrieved January 2024, from https://www.visitfinland.fi/49d063/contentassets/242f0f5314604d75b4ed84a5349dc9b3/hiilijalanjaljen-laskenta-matkailijamittarissa.pdf

OpenStreetMap contributors. (n.d.). *Extract of map data*. Retrieved June 2024, from OpenStreetMap: https://www.openstreetmap.org

Statistics Canada. (n.d.). *Small Area Estimation for Visitor Travel Survey*. Retrieved June 2024, from Statistics Canada: https://www.statcan.gc.ca/en/statistical-programs/document/5261_D3_V2

United Nations. (2010). *International Recommendations for Tourism Statistics 2008.* United Nations. Retrieved January 2024, from https://www.unwto.org/tourism-statistics/on-basic-tourism-statistics-irts-2008

Visit Finland. (2023). *Border Survey questionnaire*. Retrieved May 2024, from https://www.visitfinland.fi/49d5b2/contentassets/5ea9c50861bf4e4f969c5437ef780afa/2023-matkailijamittarin-kysymyslomake-en.pdf

Visit Finland. (2024). *Border Survey*, English. Retrieved June 2024, from Visit Finland: https://travel-trade.visitfinland.com/en/border-survey/

Visit Finland. (2024). *Matkailijamittari*, Finnish. Retrieved January 2024, from Visit Finland: https://www.visitfinland.fi/suomen-matkailudata/matkailijamittari

Visit Finland; Ministry of Economic Affairs and Employment of Finland. (2023). *Finland's tourism accounts for 2020-2021.* Visit Finland; Ministry of Economic Affairs and Employment of Finland. Retrieved January 2024, from https://www.visitfinland.fi/499076/contentassets/242f0f5314604d75b4ed84a5349dc9b3/visit-finland-matkailutilinpito-2020-2021.pdf

Visit Finland; Statistics Finland. (2022). *Matkailijamittari Perustamisvaiheen loppuraportti.* Visit Finland. Retrieved January 2024, from https://www.visitfinland.fi/48f686/contentassets/242f0f5314604d75b4ed84a5349dc9b3/matkailijamittari---perustamisvaiheen-raportti.pdf