Data Definition Summary

Fontys Travel Data Analysis

*Prepared for Team Discussion - Tuesday*

# Executive Summary

We have analyzed 5 datasets containing comprehensive travel booking information:

* **Total Records:** 53,825 rows across all datasets
* **Total Features:** 99 unique columns
* **Overall Data Quality:** 93.5% complete
* **Key Join Column:** 'boekingsnummer' (booking number) appears in multiple datasets

The datasets cover different aspects of the travel booking process and can be combined for comprehensive analysis.

# 1. Trip Overview Flight

## Dataset Overview

**Size:** 1,340 rows × 1 columns

**Completeness:** 100%

**Purpose:** Flight-related trip overview information. Contains basic trip details but appears to have formatting issues (all data in single column).

## Key Columns

| **Column Name** | **Type** | **Missing** | **Sample Values** |
| --- | --- | --- | --- |
| Export date;Reiscode;Omschrijving... | Text | 0.0% | 13-10-2025;INN;Noord-India... |

## Data Quality Issues

* ⚠️ All data compressed into single column - needs proper CSV parsing

# 2. Deelnemer Boekingen

## Dataset Overview

**Size:** 1,607 rows × 15 columns

**Completeness:** 95.2%

**Purpose:** Participant booking information. Contains customer demographics, booking details, and financial data for travelers departing in January 2025.

## Key Columns

| **Column Name** | **Type** | **Missing** | **Sample Values** |
| --- | --- | --- | --- |
| boekingsnummer | Numeric (ID) | 0.0% | 24000199, 24000201... |
| registratiedatum | Date | 0.0% | 2024-01-15... |
| startdatum | Date | 0.0% | 2025-01-06... |
| boekingsbedrag | Numeric (€) | 27.9% | €12,350.50... |
| geslacht | Categorical | 0.2% | VROUW, MAN |
| leeftijd tijdens reis | Numeric | 0.4% | 71, 61, 61... |
| nationaliteit | Categorical | 12.1% | NL, BE... |
| land | Country Code | 0.0% | NZ, IN, KE... |

## Data Quality Issues

* ⚠️ 28% missing booking amounts
* ⚠️ 12% missing nationalities

# 3. Overzicht Reizen CAX

## Dataset Overview

**Size:** 1,545 rows × 33 columns

**Completeness:** 92%

**Purpose:** Future trip planning from CAX-GV system. Comprehensive trip overview including availability, bookings, and operational details for trips departing June 2025 onwards.

## Key Columns

| **Column Name** | **Type** | **Missing** | **Sample Values** |
| --- | --- | --- | --- |
| ID | Numeric (ID) | 0.0% | 38628, 38888... |
| Departure NL | Date | 0.0% | 2024-06-01... |
| Bestemmingscode | Code | 0.0% | LAD, GEO, ALB... |
| Omschrijving | Text | 0.0% | India: Spiti & Ladakh... |
| PAX | Numeric | 0.0% | 17, 20, 16... |
| Status | Categorical | 0.0% | Boekbaar, Vol |
| MAX PAX | Numeric | 0.0% | 18, 20, 16... |
| Singles | Numeric | 0.0% | 11, 9, 8... |

## Data Quality Issues

* ⚠️ 76% missing Internet remarks
* ⚠️ 100% missing in Actie 2 column

# 4. Export Boekingen Total

## Dataset Overview

**Size:** 23,463 rows × 27 columns

**Completeness:** 91%

**Purpose:** Complete booking records export. Comprehensive booking database containing all bookings with customer, trip, and agent information through December 31, 2024.

## Key Columns

| **Column Name** | **Type** | **Missing** | **Sample Values** |
| --- | --- | --- | --- |
| BoekingsID | Numeric (ID) | 0.0% | 338370, 341202... |
| Code | Trip Code | 0.0% | NZE, INN, KTK... |
| Reiscode | Trip ID | 0.0% | NZE2401... |
| Label | Categorical | 0.0% | Puur reizen... |
| Land | Country | 0.0% | Nieuw-Zeeland... |
| Vertrekdat. | Date | 0.0% | 2024-01-01... |
| BoekingsNummer | Numeric (ID) | 0.0% | 194078... |
| Status (Boeking) | Categorical | 0.0% | Definitief... |
| *... and 2 more columns* | | | |

## Data Quality Issues

* ⚠️ 87% missing agent info
* ⚠️ 41% missing discount data

# 5. Grip Booking Freeze

## Dataset Overview

**Size:** 26,870 rows × 23 columns

**Completeness:** 97.5%

**Purpose:** Weekly booking snapshot for reporting. Comprehensive booking freeze data including revenue, customer info, and booking details for fiscal year reporting (week 251005).

## Key Columns

| **Column Name** | **Type** | **Missing** | **Sample Values** |
| --- | --- | --- | --- |
| boekingsnummer | Numeric (ID) | 0.0% | 24000196... |
| OMZET | Numeric (€) | 0.0% | €5,279, €12,298... |
| registratiedatum | Date | 0.0% | 2024-03-08... |
| startdatum | Date | 0.0% | 2025-01-06... |
| boekingsbedrag | Numeric (€) | 0.0% | €5,319.50... |
| land | Country | 10.8% | Nieuw-Zeeland... |
| boekingsstatus | Categorical | 0.0% | DEF... |
| aantal\_deelnemers | Numeric | 0.0% | 1, 2... |

## Data Quality Issues

* ⚠️ 44% missing confirmation dates
* ⚠️ 11% missing country data

# Key Findings & Recommendations

## 1. Data Completeness

* Grip Booking Freeze is the most complete dataset (97.5%)
* Trip Overview Flight has formatting issues that need resolution
* Missing booking amounts (28%) in Deelnemer Boekingen may impact revenue analysis

## 2. Dataset Integration

* **Common Join Key:** 'boekingsnummer' (booking number) found in 3 datasets
* Datasets can be combined for comprehensive customer and booking analysis
* Future trips data (CAX) is separate but complementary for planning

## 3. Data Purpose & Use Cases

* **Customer Demographics:** Use Deelnemer Boekingen for age, gender, nationality analysis
* **Revenue Analysis:** Grip Booking Freeze has complete financial data
* **Capacity Planning:** Overzicht Reizen CAX shows availability and max capacity
* **Historical Analysis:** Export Boekingen Total provides complete booking history

# Next Steps

1. **Fix Trip Overview Flight CSV parsing issue**
2. **Define primary analysis questions** (revenue trends? customer segmentation? capacity optimization?)
3. **Create merged dataset** using 'boekingsnummer' as join key
4. **Develop strategy for handling missing data** (especially booking amounts and agent info)
5. **Begin exploratory data analysis** with cleanest datasets first