Testing DTS Export to Stats New Zealand

**Accomodation (7578)** *[integrityCheck\_7578.accdb]*

|  |  |
| --- | --- |
| Step | Results |
| Data imported into MS Access environment. Dimension tables joined to main data table. Number of rows in main data table should equal to the number of rows after creating join to ensure that the dimension tables are linked to the data table | Verified total number of rows equal 157,626 |
| In Access, set up ‘<> All’ as criteria for all dimension columns. Number of rows should equal the number of rows in the structure *‘df\_base\_aggregates’.* | Verified number of rows equal 86,511 |
| Check that the group\_by combinations of all dimension levels are equal to total above | Confirmed |
| Test specific year end. Verfied that the total number of nights for year ending December 2010 was approximately equal to 47,421,565. Actual total was 474,421, 555 difference due to rounding of all the text values to zero decimal places | Confirmed |
| Spot checked various rows | Confirmed |
| Checked aggregates in df\_base\_aggregates against totals in Access: df\_base\_aggregates %>% ungroup() %>% summarise\_each(funs(sum), Total\_Visitors, Total\_Trips, Total\_Nights, Total\_Respondents) | Confirmed |
| *Check various aggregate combinations:* |  |
| *YE != All; LOS Group == All; Destination RTO == All; Accomodation Type != All*  Total number of rows confirmed as 1525. These were confirmed as unique. | Confirmed |
| *YE == All; LOS Group != All; Destination RTO != All; Accomodation Type != All*  Total number of rows confirmed as 5587. These were confirmed as unique. Totals checked. | Confirmed |
| *YE != All; LOS Group != All; Destination RTO == All; Accomodation Type == All*  Total number of rows were confirmed as 899. These were confirmed as unqiue. Totals checked. | Confirmed 530 rows |

**Trips (7579)** *[integrityCheck\_7579.accdb]*

|  |  |
| --- | --- |
| Step | Results |
| Expect that the number of observations will be 173,330 when connected to all tables. | Confirmed. |
| In Access, set up ‘<> All’ as criteria for all dimension columns. Number of rows should equal the number of rows in the structure *‘df\_base\_aggregates’.* | Yes they do. 41,645 rows. |
| Check that the group\_by combinations of all dimension levels are equal to total above | Confirmed |
| Test specific year end. Verified that for "YESep2010" that the total day and night trips were: 27,491,020 and 15,981,529 . Actual values were: 27,491,016 and 15,981,538. Discrepancies due to rounding. | Confirmed. |
| Spot check totals in df\_base\_aggregates. Spot checked the following: df\_base\_aggregates[c(10000, 20000, 30000, 40000),] | Confirmed |
| Checked aggregates in df\_base\_aggregates against totals in Access: df\_base\_aggregates %>% ungroup() %>% summarise\_each(funs(sum), Total\_Visitors, Total\_Trips, Total\_Nights, Total\_Respondents) | Confirmed. Some differences due to rounding. |
| *Check various aggregate combinations:* |  |
| YE =! All, DestinationRTO == All, TripType != All, POV == All, POV\_Group != All | Confirmed |
| YE == All, DestinationRTO != All, TripType == All, POV != All, POV\_Group == All | Confirmed 701 rows |
| YE != All, DestinationRTO == All, TripType != All, POV == All, POV\_Group == All | Confirmed 106 rows |

**Spend (7580)** *[integrityCheck\_7580.accdb]*

|  |  |
| --- | --- |
| Step | Results |
| Data imported into MS Access environment. Dimension tables joined to main data table. Number of rows in main data table should equal to the number of rows after creating join to ensure that the dimension tables are linked to the data table | Verified total number of rows equal to 1,502,092 |
| In Access, set up ‘<> All’ as criteria for all dimension columns. Number of rows should equal the number of rows in the structure *‘df\_base\_aggregates’.* | Verified number of rows equal to 313, 577 |
| Check that the group\_by combinations of all dimension levels are equal to total above | Confirmed |
| Confirm that totals for Year Ending September 2010 are equal. Confirmed totals for the POV\_Group reconcile. Confirmed totals for Day Trip and Overnight Trip reconcile. Confirmed that total number of trips approximately (due to rounding) reconcile to SAS reports as documented in R script (see lines 150 (approx) | Confirmed. |
| Spot checked the following: df\_base\_aggregates[c(100000, 200000, 300000),] | Confirmed totals match |
| Checked aggregates in df\_base\_aggregates against totals in Access:  *sum(df\_base\_aggregates$Expenditure)* |  |
| *Check various aggregate combinations:* |  |
| df\_consolidated %>% filter(Year\_Ending == "All" & Destination\_RTO != "All" &  Trip\_Type != "All" & POV == "All" & POV\_Group == "All" & Spend\_Type == "All" ) %>%  summarise(count = n()) | This is 66 rows |
| df\_consolidated %>% filter(Year\_Ending != "All" & Destination\_RTO == "All" &  trip\_Type == "All" & POV != "All" & POV\_Group == "All" & Spend\_Type != "All" ) %>%  summarise(count = n()) | 9475 rows |
| df\_consolidated %>% filter(Year\_Ending == "All" & Destination\_RTO == "All" &  Trip\_Type != "All" & POV == "All" & POV\_Group == "All" & Spend\_Type != "All" ) %>%  summarise(count = n()) | 15 rows |

**Spend (7581)** *[integrityCheck\_7581.accdb]*

|  |  |
| --- | --- |
| Step | Results |
| Data imported into MS Access environment. Dimension tables joined to main data table. Number of rows in main data table should equal to the number of rows after creating join to ensure that the dimension tables are linked to the data table | All rows joined. 3247 rows. |
| In Access, set up ‘<> All’ as criteria for all dimension columns. Number of rows should equal the number of rows in the structure *‘df\_base\_aggregates’.* | Confirmed number of rows are 3127 |
| Check that the group\_by combinations of all dimension levels are equal to total above | Confirmed. |
| Confirmed that total matches for december 2000. Small differences apparent due to rounding.  df\_base\_aggregates %>% filter(YE == "YEDec2000") %>% summarise(total = sum(Total\_Activities))  Total was 122499731.1 | Confirmed |
| Spot checked the following. df\_base\_aggregates[c(900, 2700),] | Confirmed |
| Checked the the following sum(df\_base\_aggregates$Total\_Activities) | Confimed (differences due to rounding) |
| *Check various aggregate combinations:* Confirmed that there are 67 rows when YE == “All” Confimed that there are 54 rows when ActivityGroup is set to “All” Spot checked various values. | Confirmed. |
|  |  |

Code: 7578

Title: Accommodation

Number rows: 157,627

Number dimensions: 4

Number measures: 4

Approx. size (zipped): 1.7 Mb

Number of files: 12

Code: 7579

Title: Trips

Number rows: 173,331

Number dimensions: 5

Number measures: 3

Approx. size (zipped): 1.6 Mb

Number of files: 14

Code: 7580

Title: Spend

Number rows: 1,502,093

Number dimensions: 6

Number measures: 1

Approx. size (zipped): 9.6 Mb

Number of files: 16

Code: 7581

Title: Activity

Number rows: 3,248

Number dimensions: 2

Number measures: 1

Approx. size (zipped): 19 Kb

Number of files: 8