**Background**

This is the analysis we normally do for travel client but is also applicable for hotel booking and even ecommerce site.

There are 3 phases in the customer’s journey:

* **Dream / Search:** When users start visiting the site to browse for flight without making a booking
* **Book:** When users come to the site and make a booking
* **Fly:**  When users take their departure flight

In airlines, it is important to clear **“load factor”** (i.e: Share of unsold seats on the plane) hence the earlier they can sell the seat the better. Traditionally, the commercial team (or also called revenue team) controls the inventory and make decision based on **sales data** to plan out promotion, fare and deals. And marketing team is normally the one the executors.

What the they don’t have or make use of are the users’ interaction data prior to make a booking, which are strong indicators of buying intention to factor that in their strategy to clear “load factor” faster.

This analysis deep dive into the full user behaviour from research phase to flying.

Key factors that normally impact this behaviour, typically for airlines are:

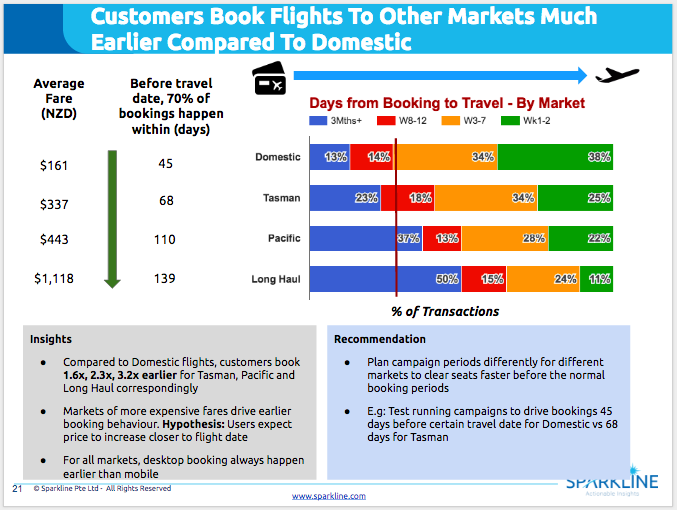
* Price (Fares): Higher fare, longer buying cycle
* Market Type: (e.g: Long haul, short haul, domestic) The longer the route, the higher the price
* Travel Period: Peak season is expected to be have higher demand hence higher price
* Device: Mobile is normally used more for research and desktop is more for making the final booking

Please refer to the docs below for some actual analysis/output

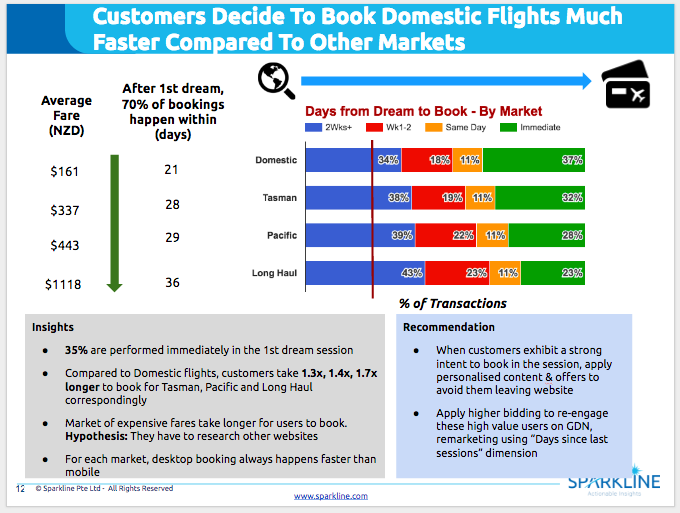
* [Air Asia](https://docs.google.com/presentation/d/1ndlnoQJE39k7bxUcgmeXbbWJL9hS1K_qDHcrBpBIsdo/edit#slide=id.g104bf5c365_0_115)
* [Air NZ](https://docs.google.com/presentation/d/1RPDdWw81e1RA-kch0J2d4D0gMUabe7rMjreoz4my0j4/edit#slide=id.g1d01d710cb_0_0) (I will use Air NZ as the example in the instruction below)

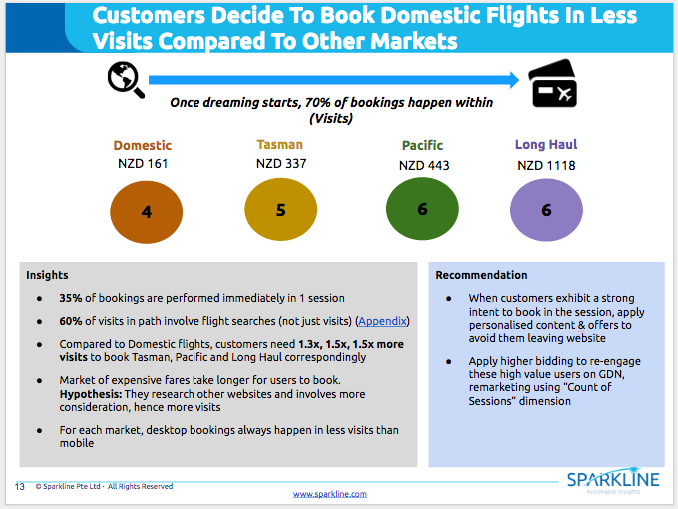
**Analysis Output**

**Insight: No. of days / interactions from Book-to-Fly**

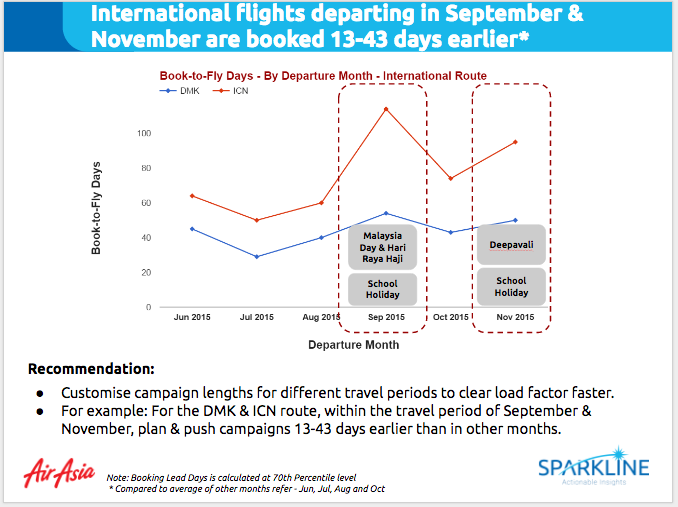
****

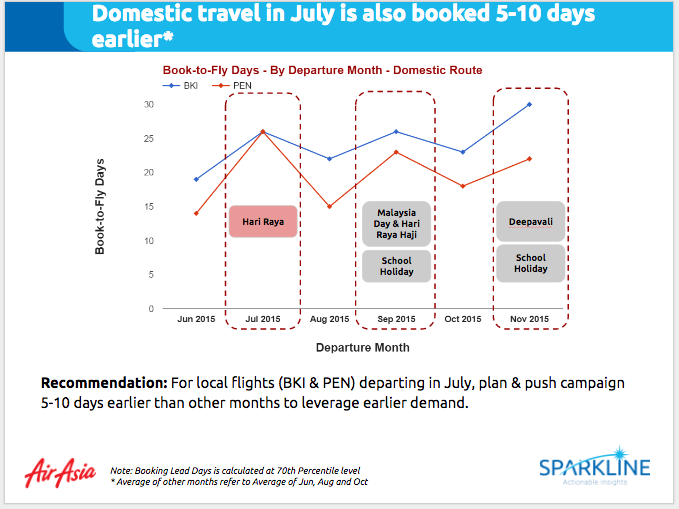
**Insight: No. of days / interactions from Dream-to-Book**

****

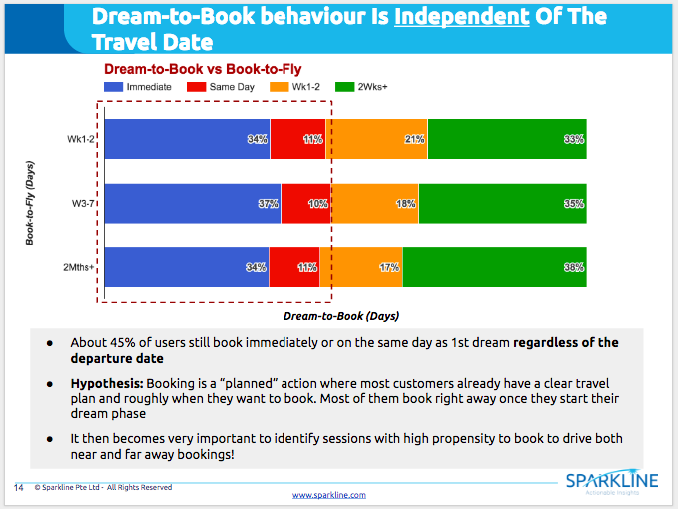
****

**Insight: Different Dream to Book for Different Travel Period (for different market)**

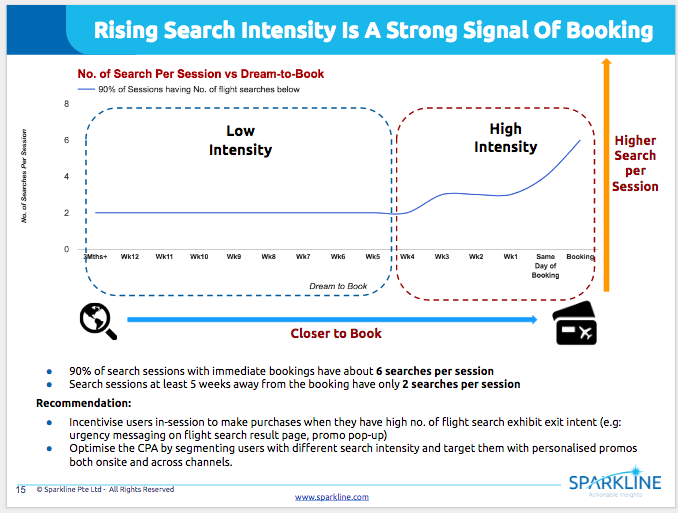
****

****

**Insight: Relationship between Dream-to-Book vs Book-to-Fly**

****

**Insight: Search intensity is key indicator for booking intention**

****

**Requirement:**

* Access to BigQuery or detailed clickstream and transactional data
* Discuss clearly with clients to scope out, what are the relevant segmentation that they want to do. For example for Air NZ, we focus on Device Type, Market Type and Search Lead Days. This is really important as it will decide what are the additional informations that you need to extract from BigQuery along with your transaction information.

**Analysis Execution:**

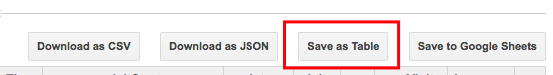
* Please treat this document as a guideline rather than a manual, depending on your specific analysis requirement, please adjust accordingly
* All sample codes can be found [here](https://drive.google.com/open?id=0B_6oGZcbKKqhS2xBWDN1Z0RuMDA)

1. Decide time period

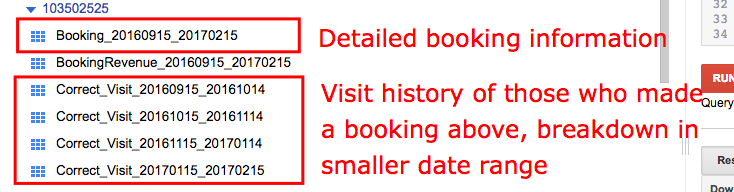
* Time period 1 [TP1]: To extract booking/transaction information [X to Y]
* Time period 2 [TP2]: To extract visit history of the users who make a booking during time period 1 above. If you have long history of data just use as long as you can, otherwise, pick the date range where the data is most clean to ensure accurate analysis. 90-120 days are normally good [X-90 or X-120 to Y]

2. Extract detailed visit history during [TP1] of users who make a booking during [TP2]:

* [Refer to this sample BQ code](https://drive.google.com/open?id=0B_6oGZcbKKqhNXNYOHV4MndScDA)
* If the look back window that you use is too large (1 year etc), you should keep the same date range of [TP2] and breakdown [TP1] to the smaller date range that you can run
* Once you can run it successfully, just “Save as table” as below



* For example: *In the example below, there is a separate Booking Revenue table because I forget to extract the revenue figure at first so I extracted that separately afterward so that I didn’t have to redo everything*



3. Extract transaction information of the booking happened during [TP2]

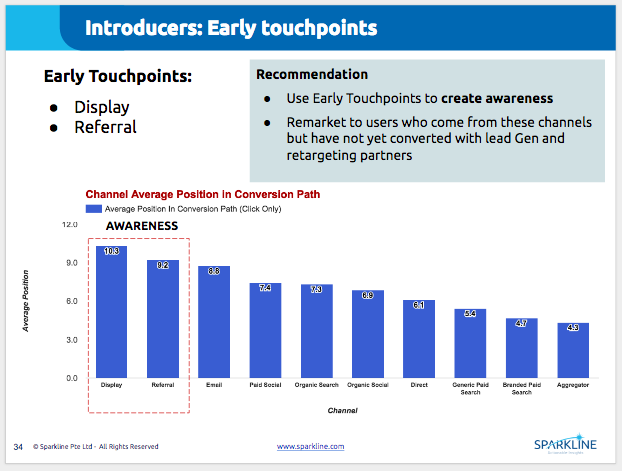
* [Refer to this sample BQ code](https://drive.google.com/open?id=0B_6oGZcbKKqhekVNZ0RDZFVqQ1U)
* Note: The information used to segment the booking (e.g: Device, Market …) will be extracted along in this query

4. Use R to manipulate these 2 data outputs to produce all the insights above

* [Sample R code with detailed comments](https://drive.google.com/open?id=0B_6oGZcbKKqhdlUyUDMteW44NVU)
* [Refer to this folder for all the output to google sheet for visualization](https://drive.google.com/open?id=0B_6oGZcbKKqhT0VCZjg4WEVnbVk)

**CUSTOM ATTRIBUTION**

**Insight: Average position in path**

****

With the same methodology of extracting data above, you can also do the custom attribution as you can see in the last portion of the R code example above.

You can then rebuild any custom attribution you like (time decay, position based etc) by assigning the relevant credits based on the position in the path.

For data-driven model, the idea behind is the markov model, these are some good articles that you can use to try out ourselves, with the detailed codes:

* <http://www.lunametrics.com/blog/2016/06/30/marketing-channel-attribution-markov-models-r/>
* <http://stuifbergen.com/2016/11/conversion-attribution-markov-model-r/>
* <https://cran.r-project.org/web/packages/ChannelAttribution/ChannelAttribution.pdf>

**Benefits of this vs normal GA interface:**

* No limit on look back window of 90 days like GA
* You can conduct this for **any goals** that you like. In this case, we do flight booking specifically, but you can also do for sign up, add-to-cart etc …
* You can segment the attribution output by any relevant dimensions. In this example above, you can segment the insight by device, market type ... For the case of e-commerce, you can segment by product category, price etc ...
* You can override any channel tagging problem that they have in the past so that you don’t have to wait for them to fix it, wait few months to collect data to re-analyse
* In case (many clients) they have huge self-referral problems from payment gateway (i.e: it steals all the transactions of the original traffic source), then with BigQuery you can retrieve this with some manipulation (Idea: if the source & medium of a transaction is that payment gateway then use the source & medium of the immediately previous session)

**Downside**

* With BQ and GA360, we don’t have the impression interaction (which is quite significant for Display traffic) and only click data.
* Idea: Can ask for Data Transfer of DoubleClick in BigQuery to merge to retrieve impression data

