

# Meeting minutes

- Team SkyTrek

Date (dd.mm.yyyy)	01.03.2016
Attendee(s)	Jedaiah, Aseem, Huy, Ms. Antoinette (Skyscanner Content Manager)
Absentee(s)	-

---

## Contents

---

Item	Discussion point	Remark
1	<p>Report of Findings Categorization of Content Themes</p> <p>Correlation Analysis</p> <p>Regression</p>	<p>We manually classified each article into the seven provided content themes of Skyscanner.</p> <p>There is a large number of articles without clear classifications.</p> <p>We suggested Skyscanner adding 2 more content themes: Food and Activity/Topic discussion, due to their significant number of matching articles.</p> <p>Only done on data with organic attributes (i.e. not paid content)</p> <p>Articles with longer length, more videos, links and images have higher organic page view.</p> <p>High exit rate and bounce rate correlate with high time on page.</p> <p>There is a negative correlation between page view and time on page.</p> <ul style="list-style-type: none"><li>➔ Many people go to the article but do not read it</li><li>➔ We suggested Skyscanner decide on a specific measure, either page view or time on page</li></ul> <p>Regression result doesn't show much insights, not enough to predict trends or a definite deciding factor for a good article.</p>
2.	<p>Skyscanner's concern about growth from different source/medium</p>	<p>In the long run, Skyscanner is interested in articles that perform well across multiple sources/ mediums.</p> <p>Recommendation</p> <ul style="list-style-type: none"><li>- identify which source/medium work the best with which content themes</li><li>- identify which source/medium work the best with which measures (ATOP, UPV)</li></ul>

		<p>Threshold / cutoff points for well performed articles</p> <ul style="list-style-type: none"><li>- UPV (organic): 400</li><li>- ATOP (organic): 90</li></ul>
3.	Title Analysis	<p>We suggested the analysis of content themes derived from article titles</p> <p>Skyscanner suggested separating between article about local and oversea destination.</p>

## Actions

Task	Assigned to	Status
Add article titles to dataset	Huy	Created
Update dataset with content theme allocation for each URL	Jedaiah	Created
Analyze text mining clusters for titles	Jedaiah	Created
Run regression analysis again	Aseem	Created
Explore details of incorporating categorical and numerical data into regression model	Huy	Ongoing
Get Skyscanner NDA to Prof Kam	Aseem	Ongoing
Wiki update for week 8	Aseem & Jedaiah	Done
Manual classification of content types based on heavy weightage terms generated from k means model	Jedaiah	Done
Book consultation slot with Prof Kam for week 8 Tuesday this Sunday <ol style="list-style-type: none"> <li>To clarify with Prof Kam on issues mentioned in proposal feedback</li> </ol> How should we account for the NDA since the final conference paper would eventually be published online?	Jedaiah	Done
Complete wiki updates for week 7 <ol style="list-style-type: none"> <li>Minutes clean up and upload</li> </ol> Updated Gantt chart	Huy	Done
Complete data pull of 7 datasets and get them verified by Aseem Remarks: Huy will be out of town for week 8 and must ensure that pulled dataset is accurate and ready for analysis in week 8	Huy	Done
Pull Google Analytics Data	Aseem	Done
De-aggregate shares count for each URL	Jedaiah & Huy	Done
Build regression model based on current dataset	Aseem	Done
Come up with list of possible title categories to clear with sponsor	All	Done
Clarify with Prof Kam: Is there a way to dynamically assign new articles to the relevant clusters based on the trained model?	All	Done
Characteristics to scrape from article: <ol style="list-style-type: none"> <li>Number of words (remove stopwords)</li> <li>Number of links</li> <li>Images</li> <li>Videos</li> <li>Number of shares of article (by Jed)</li> </ol>	Huy	Done
Retrieve URL share count	Jedaiah	Done
Put topic modelling and google trends research and implementation into Gantt chart timeline	Jedaiah	Done
Input meeting with client in gantt chart for: <ol style="list-style-type: none"> <li>25th Jan</li> <li>2nd Feb</li> <li>11th Feb</li> </ol>	Jedaiah	Done
Create flow chart documenting process how skyscanner team plans for the next quarter's content postings. To send it to client for verification	Aseem	Done
Test Google API public holiday pulling	Huy	Done

<p>Create a mockup for the following visualization:</p> <ul style="list-style-type: none"> <li>○ To plot Unique page views views (count) vs Week of article publishing</li> <li>○ Clicking this particular bar of interest (week has high count) would drill into the next bar chart (Number of unique page views, vs identifying URL)</li> <li>○ Further clicking on the URL would reveal characteristics of the article (e.g. number of images, videos, etc, bounce/exit rate, etc)</li> </ul>	Jedaiah	Done
<p>Create a mockup for the following visualization:</p> <ul style="list-style-type: none"> <li>○ Identify key features of popular posts (MLR against unique page views)</li> </ul>	Huy	Done
<p>Create a mockup for the following visualization:</p> <ul style="list-style-type: none"> <li>○ Where do they come from (On the whole country news page level, where do the traffic come from)</li> <li>○ How long do they spend at the page (average time on page)</li> <li>○ Do they visit other pages? (bounce rate vs exit rate)</li> <li>○ How effective are the recommended pages? (If GA does not track the unique page views, then this analysis is not going to be implemented)</li> <li>○ [Just to prove a logical point] Correlation test between number of shares vs unique page views</li> </ul>	Jedaiah	Done
<p>Deal with proposal segments:</p> <ol style="list-style-type: none"> <li>1. Sponsor and Background Information</li> <li>2. Motivation</li> <li>3. Objectives</li> <li>4. Data</li> <li>5. Scope of Work</li> </ol>	Aseem	Done
<p>Questions for Prof Kam:</p> <ol style="list-style-type: none"> <li>1. Ask Kam about requirement to post sample date. Concern: data confidentiality</li> <li>2. What is required for scope of work? (different from methodology)</li> </ol>	All	Done

---

*The minutes were prepared by: Jedaiah Tan*