

# ANLY482 MEETING MINUTES WITH SPONSOR (07 March 2018)

<b>Date:</b>	7 March 2018
<b>Time:</b>	09:00-11:00
<b>Venue:</b>	SMU SIS Meeting Room 4.1
<b>Attendees:</b>	Team: Ruiyan, Qian, Nicholas, Sponsor: Prof Kam
<b>Agenda:</b>	Exploratory Findings for Kiva Dataset

S/N	Things Discussed/Done	Remark
-----	-----------------------	--------

1.	Discussion of findings	<ul style="list-style-type: none"> <li>• EDA to be done on a global level, followed by country. Time frame and time-based analysis should be conducted, over the 4 years of data, and both percentages and absolute values should be included in charts</li> <li>• Study how the loan amount and quantity changes over time. Was Philippines always high for the last 4 years, or was Kenya higher in any of the years?</li> <li>• Ensure sponsor is able to view data of the different countries in one go, and also show the trend of repayment not just the loan amount.</li> <li>• Although a single loan may have both male and female borrowers, consider the main person responsible for the loan and use his gender for analysis and being representative of the loan road</li> <li>• Activity type is too generic – similar activities should be under the same category, such as pigs, livestock and poultry falling under ‘animal sales’</li> </ul>
----	------------------------	--

		<ul style="list-style-type: none"> <li>• A dashboard is need with the flexibility to zoom into any country of choice, not just Philippines.</li> <li>• Reorganise charts in the form of a dashboard to get a 360° overview of loan details. Include temporal changes over time, and peg into a Tableau Workbook.</li> <li>• For data cleaning, state issues identified actions taken and additional useful comments (such as number of cases). Validate columns first, clarify with sponsor during the meetings before deciding on removing columns.</li> <li>• Tags are included in data to make the loan easier to find during a website search. JMP has text analysis module that allows frequency count quickly. From managerial aspect, it is not important (to be revisited at a later stage). For sponsor, only the date of the record is important. Time of day not important (unlike e-marketing/e-commerce data where timing of purchase matters). Currency might be important because of currency fluctuation rates. If not in USD, local currencies are subjected to fluctuations and inflation.</li> <li>• For records with missing data, some have the location name. Fill in the latitude and longitude based on the location.</li> <li>• For loan_theme_ids, handle as missing value when joining the table as loan id is valuable. Loan theme by region contains redundant columns as they mean the same thing, some have more records than others as they include both new geo and old geo. Choose to only retain latitude longitude to keep only meaningful columns.</li> <li>• Go to column, recode function to make it same as GADM, match with loan data. Remove the 3<sup>rd</sup> level region from analysis, only first two levels (region and city/municipality) are sufficient. Save the recoded fields as new columns.</li> </ul>
--	--	---

		<p>Subsequently, work on statistical analytical methods. Some regions are in below municipality levels, map them back up to municipality levels.</p> <ul style="list-style-type: none"> <li>• Use UML diagram when explaining how the files are cleaned and joined (primary key/unique id, secondary key).</li> <li>• Focus on a particular methodology (geospatial analysis first, then text analysis if time permits). Find out the relationship between funded amount and time difference/duration of funding, leave as these 2 columns first instead of diving into rate of funding. For tree maps, organise variables hierarchically (sector, then activity). Among agriculture, analyse how the different activities are divided. Size should represent the number of loans, colour should represent the total amount. Filter should be country, region, year for dashboard. Replace filter to country such that we can get the tree map for each individual country straight away. Tree maps can also be based on continent, countries, regions, filter by sector. Analyse how each sector is distributed around the world. Line or horizon graph (which shows comparison of different sector) can be used for temporal patterns, slope graph which allows how relative positions change. Filter by country, show trends for top 10,20,50 sectors and how the distribution and loan amount change over time.</li> <li>• Future agenda to include recap of past meeting and plan for upcoming week(s)</li> </ul>
2.	Follow-up actions	<ul style="list-style-type: none"> <li>• Retrieve administrative region data from GADM for top few countries</li> <li>• Prepare interim report</li> </ul>

**Item Due (Team) / Actions**

Deadline: End of week.

1. Interim report to be handed in by 11/03/2018, Sunday.