

# Master of Data Science - University of British Columbia Okanagan

Minutes for May 21, 2020 - 1:45 PM - 2:45 PM PST

**Present:** Sofia Bahmutsky, Ngan Lyle, Kaitlyn Hobbs (Minutes),

**Absent:** Shreeram Murali

## Agenda

1. Update teammates on progress.

## Proximity Data Wrangling\*

\* *Confirm methodology with Bruno and Marian*

**Note:** *in\_db* is a binary column that represents if there is the proximity item *within* that dissemination block.

- Rows with "F" indicates that data is too unreliable - leave for GIS, drop for PCA
- Rows with "." indicates that data is not available (NA) - convert to np.nan within pandas.
- Rows with blanks - ?
- **Fixing Data Types:**
  - In Python, convert raw PMD-en data entries with "." to np.nan.
  - Export to Excel
  - Drop into QGIS
  - Click Field Calculator (top bar menu) > Create a new field > Create virtual field > Output field name; Output field type: Decimal number (double) > **to\_real (prox\_column\_name)**

## Action Items

-