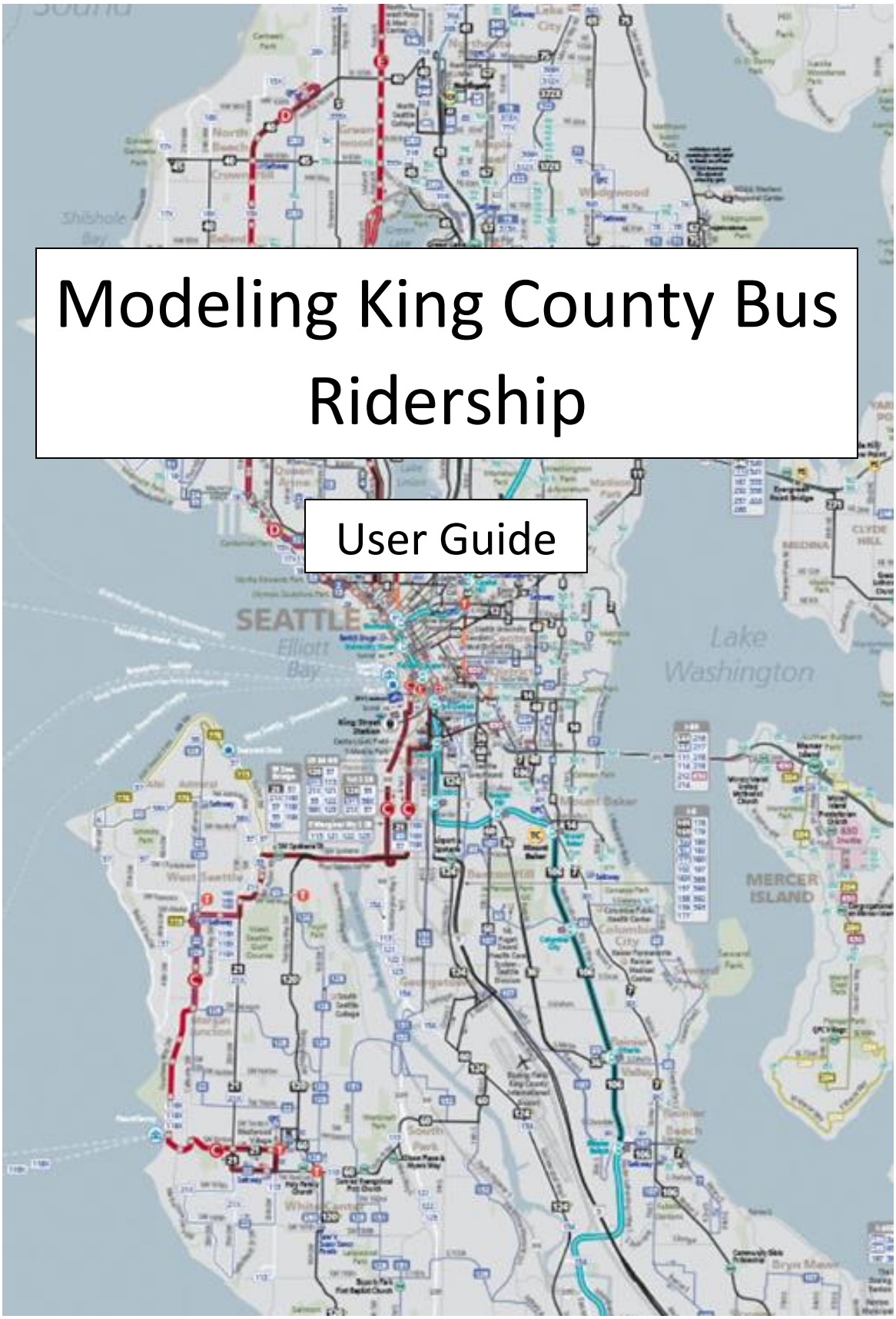

Modeling King County Bus Ridership

User Guide

Setup Environment:

1. Clone the repository:

   >> git clone https://github.com/jacobw125/uw-msds-trac-capstone.git

2. Create a new python environmet using the command:

   >> conda env create UWTRAC

3. Activate UWTRAC by using the command:

   >> conda activate UWTRAC

4. Install the required python packages using:

   >> pip install –r requirements.txt

Creating Training Sets:

Regenerate the training sets from 2019 winter and/or summer APC survey period:

Summer:

```
>> python3 pipeline/01_filter_apc.py Summer_APC_File True

>> python3 pipeline/02_filter_orca.py Summer_ORCA_FILE True

>> python3 pipeline/03_agg_orca.py True

>> python3 pipeline/04_merge.py True

>> python3 pipeline/05_create_training.py S
```

Winter:

```
>> python3 pipeline/01_filter_apc.py Winter_APC_File False

>> python3 pipeline/02_filter_orca.py Winter_ORCA_FILE False

>> python3 pipeline/03_agg_orca.py False

>> python3 pipeline/04_merge.py False

>> python3 pipeline/05_create_training.py W
```

Combined:

Run steps 1-4 for for Summer and Winter, then:

```
>> python3 pipeline/05_create_training.py B
```

Generate new training sets from different APC survey period:

1. **Update constants in pipeline/constants.py:**
   a. Check column headers of new files for ORCA (ORCA_SUMMER_COLUMNS, ORCA_WINTER_COLUMNS) and APC (APC_COLUMNS) match constants. If not, either update constants.py or relabel your file.

b. SUMMER_DAYS or WINTER_DAYS to desired date range. Possibly excluding extreme weather (i.e. snowstorm) that could potentially negatively impact your model.

2. **Run pipeline:**

Note:

For steps 1-4, the True/False parameter refers to is this is a summer survey data set, directing the programs to use SUMMER constants instead of WINTER constants.

For step 5, S means use SUMMER constants and W means use WINTER. In order to use the parameter B you much run steps 1-4 on both a WINTER and SUMMER set to create the COMBINED training set.

>> python3 pipeline/01_filter_apc.py NEW_APC_FILE_PATH.csv [True/False]

>> python3 pipeline/02_filter_orca.py NEW_ORCA_FILE_PATH.csv [True/False]

>> python3 pipeline/03_agg_orca.py [True/False]

>> python3 pipeline/04_merge.py [True/False]

>> python3 pipeline/05_create_training.py [S/W/B]

Add weather data to the model:

1. **Get data:**
2. **Update constants in pipeline/constants.py:**
3. **Uncomment code**
4. **Run steps 1-5 in pipeline**

Training Model:

Running the Model: