**Methodology**

**Data:**

In conducting our analysis, we fused a range of datasets (summarized in Table #). Our primary metric for judging ridership came from Automatic Passenger Counter (APC) data, which shows boardings and alightings at every stop on a bus route. We then used bus stop coordinates to aggregate ridership to the census tract level, allowing us to use a variety of sociodemographic, economic, and spatial factors in our analysis.

Table #: Datasets fused during ridership analysis

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Information** | **Purpose** | **Source** |
| APC Counts | Bus boardings and alightings per stop | Assess ridership |  |
| Bus Stops | Every bus stop location (i.e., coordinates) in King County |  |  |
| Census Tracts | ? | ? |  |
| ACS Data | Income, education, race, age, gender, share of transport modes to work, etc. | Merging sociodemographic information |  |
| Employment | ? | Merging employment figures |  |
| POIs | Food and medical facilities, school sites, public parcels, etc. | Merging distribution of POIs |  |
| Land Use | Share of land belonging to commercial, residential, or industrial zoning | Merging land use proportions |  |
| Population Density | Did we have this? |  |  |

**Study Framework**

We partitioned the recent history of Seattle bus ridership into the following four phases:

1. Pre-Covid: March 2019 to December 2019 (Baseline #1)
2. Covid (peak era): March 2020 to December 2020 (Baseline #2)
3. Covid (Transition): March 2021 to December 2021
4. Recovery: March 2022 to December 2022

We chose these dates based on a set of important events that significantly influenced ridership. On January 21st, 2020 the U.S. had its first confirmed COVID-19 case in Snohomish County, WA. This was followed by the first COVID-19-related death in the U.S., which happened in King County. As visible in Figure #, bus ridership dropped significantly in those two months. On February 1st, 2021, two mass COVID-19 vaccination sites opened in King County and a significant amount of public aid was disbursed in the national vaccination effort. This marked the start of the transition period. Then, in November 2021, the world was made aware of the new Omicron variant, subsequently leading to a pronounced increase in cases and deaths and a noticeable decrease in ridership. By March 2022, however, with declining case rates, increasing vaccination rates, and the lifting of mask mandates, western U.S. states gradually returned to pre-pandemic operations and beliefs, and hence we deem the ensuing period the recovery phase.

Assessing ridership recovery can be interpreted in one of two ways: (1) What percent of pre-pandemic ridership levels have we recovered (i.e., 100% corresponding to a full recovery) and (2) how much has ridership rebounded since rock bottom (i.e., since peak COVID-19 period)? We analyze ridership recovery with respect to both of these baselines, highlighting the differences and similarities.

Chart, scatter chart

Description automatically generated

Figure #: Average daily bus ridership patterns in King County, 2019-2023

**Preprocessing**:

We began preprocessing by grouping and aggregating APC counts at different spatiotemporal levels (i.e., daily, weekly, monthly, stop level, and stop-daily) to conduct descriptive analyses. The count data from each study period were grouped and aggregated to the daily-stop level, and then divided into weekday and weekend datasets for further analysis. For each stop, we also calculated the number of days that data was collected. This number varied significantly for different stops and different years. This is probably due to the issues and limitations of APC data.

To get average daily ridership at each stop during a given period, we divided the total ridership at the stop in that period by the number of days in the period:

where is the row-vector of average daily ridership at bus stop during period , is the total ridership for each stop during the given period, and is the number of days in the period. We then aggregate these values to the census tract level based on the geographical location of the stop. For our analysis on ridership rebound (i.e., with respect to peak-COVID baseline), we estimated the ridership change rate for each census tract, separating weekdays and weekends. We repeated this process for our analysis on ridership recovery (i.e., with respect to pre-COVID baseline).

**Processing**

Descriptive analysis:

ANOVA was used. Temporal trends are plotted. (\*\* we will probably add more descriptive analysis\*\*)

Spatial analysis:

The ridership change rate of closeness and recovery for weekends and weekdays was plotted on the map using R. (\*\*I will also plot the stop locations\*\*), We also have food facilities, medical facilities, POI, and STOP number plots.

Linear regression:

Linear regression was used to find the contributing factors to uneven ridership change among census tracts.