Analysis Report on Publibike Dataset

November 17, 2019

1 Data Description:

We merged the following data:

- Publibike: total number of Publibike rides by hour
- Swisscom: total number of Swisscom commuters in Lugano by hour
- TPL: total number of total running buses by hour

2 Substitution Effect btw Buses and Publibikes

2.1 All Sample

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \tag{1}$$

where y_i represents the number of rides per hour, and x_i represents the number of Swisscom commuters per hour.

Table 1: All sample

	Dependent variable:
	# rides per hour
# Swisscom commuters per hour	0.005^{***}
	(0.001)
Constant	24.172^{***}
	(1.070)
Observations	768
R^2	0.072
Adjusted R^2	0.071
Residual Std. Error	18.833 (df = 766)
F Statistic	59.793^{***} (df = 1; 766)
Note:	*p<0.1; **p<0.05; ***p<0.01

2.2 By bus coverage

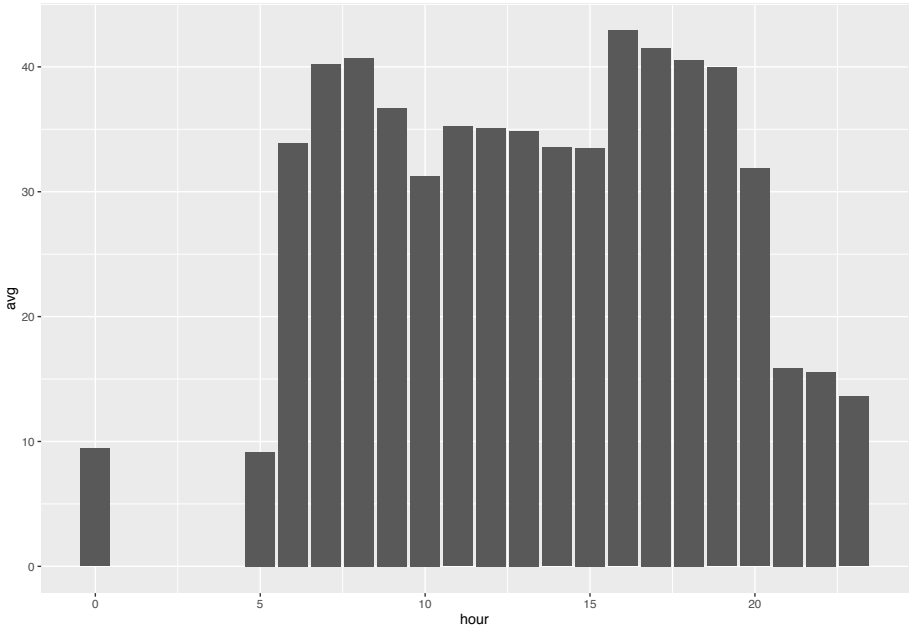

Figure 1: Histogram of number of running buses by hour

$$y_i(S) = \beta_1 + \beta_2 x_i(S) I_D + \varepsilon_i \tag{2}$$

where $S = D, E, N$.

- D : day time 6am-8pm, enough buses on the road
- E : evening time 9pm-0am and 5am, a few buses on the road
- N : night time 1am-4am, no buses on the road

Table 2: Subsititution effect by bus coverage

	Dependent variable: # Publibike rides per hour		
	6am - 8pm	9pm - 0am and 5am	1am - 4am
# Swisscom commuters	0.003***	0.088***	0.113**
	(0.001)	(0.013)	(0.043)
Constant	28.600***	7.525**	3.666*
	(1.657)	(3.729)	(2.150)
Observations	521	174	73
R^2	0.022	0.212	0.089
Adjusted R^2	0.020	0.207	0.076
Residual Std. Error	16.594 (df = 519)	21.521 (df = 172)	8.434 (df = 71)
F Statistic	11.759*** (df = 1; 519)	46.293*** (df = 1; 172)	6.902** (df = 1; 71)
Note:			*p<0.1; **p<0.05; ***p<0.01

2.3 Conclusions and suggestions

1. A lower ratio of commuters will use Publibikes when there are more buses available
2. From 9pm to 4am, around 10% of commuters use Publibikes, which is way higher than day time, 0.3%
3. During the evening time, the R square of the model reaches 21.2%, which means 20% of the changes in demands for Publibikes are driven by commuters.
4. Publibike company should consider pushing ads designed for commuters during evening time. (Precise marketing)

3 User Type Impact on the Usage of Publibike

Causality Issue: Both the usage of Publibikes and buses will be driven by the commuting demands during peak hours. We need to control for that by gathering the residuals first:

$$e_i^y = y_i - \hat{\beta}_1 + \hat{\beta}_2 z_i \tag{3}$$
$$e_i^x = x_i - \hat{\gamma}_1 + \hat{\gamma}_2 z_i \tag{4}$$

where z_i is the number of buses that are on the road. Then we run the following regression:

$$e_i^y(T) = \theta_0 + \theta_1 e_i^x(T) + \varepsilon_i, \tag{5}$$

where T = 'summer' or 'not summer'.

Table 3:

	Dependent variable:	
	Residuals of (# number of Publibike rides)	
	Before summer holiday	Summer holiday
Residuals of (# number of Swisscom commuters)	0.003**	0.0005
	(0.001)	(0.002)
Constant	−0.000	−0.000
	(0.923)	(1.062)
Observations	237	284
R^2	0.028	0.0003
Adjusted R^2	0.024	−0.003
Residual Std. Error	14.216 (df = 235)	17.890 (df = 282)
F Statistic	6.711** (df = 1; 235)	0.094 (df = 1; 282)
Note:		*p<0.1; **p<0.05; ***p<0.01

3.1 Conclusions and suggestions

1. Students are the major users of Publibikes
2. Employees don't ride Publibikes more often even if there are more bikes available during the summer holiday.
3. To promote sustainability, employers should encourage employees use Publibikes more often especially during the summer holiday.