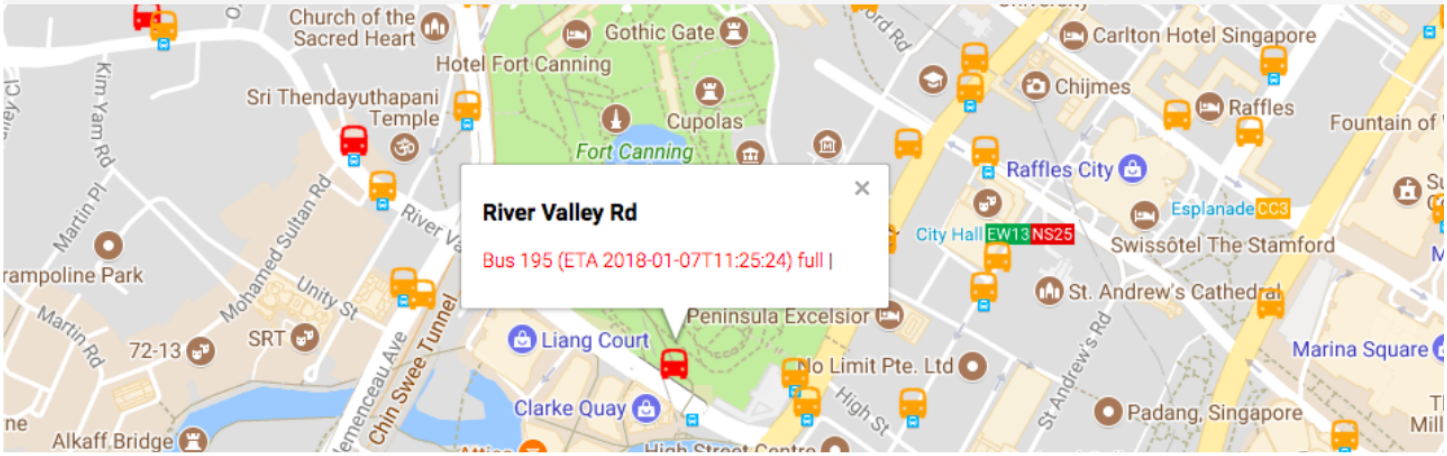


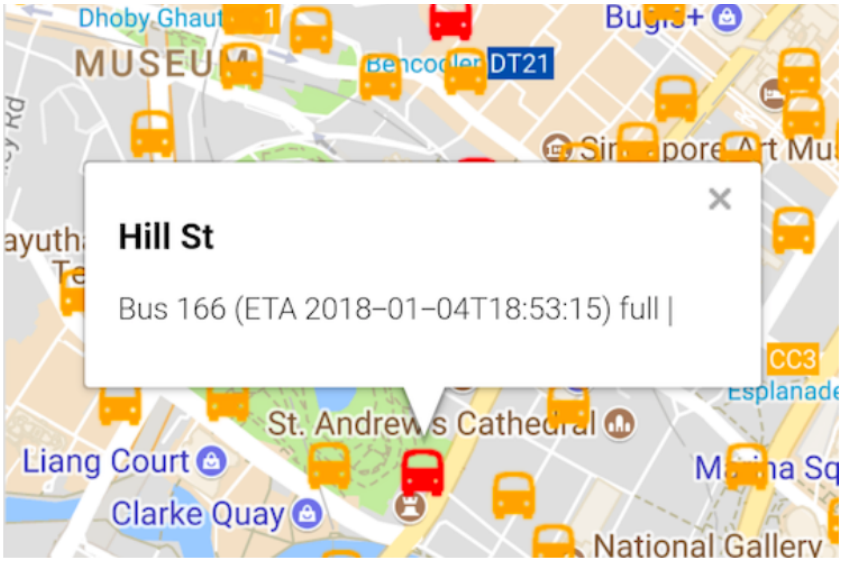
Making sense of Open Data & APIs: Singapore's busy buses

Published on January 7, 2018

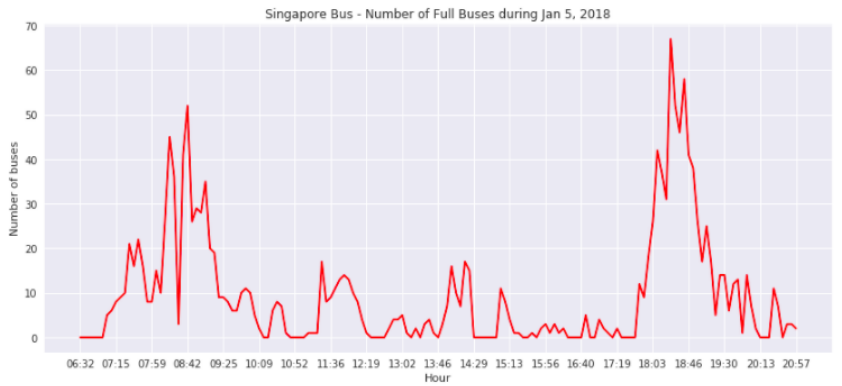
I'm sitting at Starbucks with my laptop, accessing internal systems of a major bank. Not too long ago, this was considered hacking and a punishable offence - today we live in a world of APIs and Open Data. Let's work with some of that material and have fun!

Did you know that the International Space Station (ISS) has an API? You can try it in the Did you know that the International Space Station (ISS) has an API? You can try it in the browser to get the current location or to find out when exactly it's flying over Singapore next time. For all the non-geeks in the audience: congratulations, if you clicked on these links you can now say you've triggered an API call, and you might even be able to make sense of the data snippets you see in the browser. The ISS API is just one of many examples, an increasing number of companies and institutions are making some of their data available for everyone to consume and build mashups.

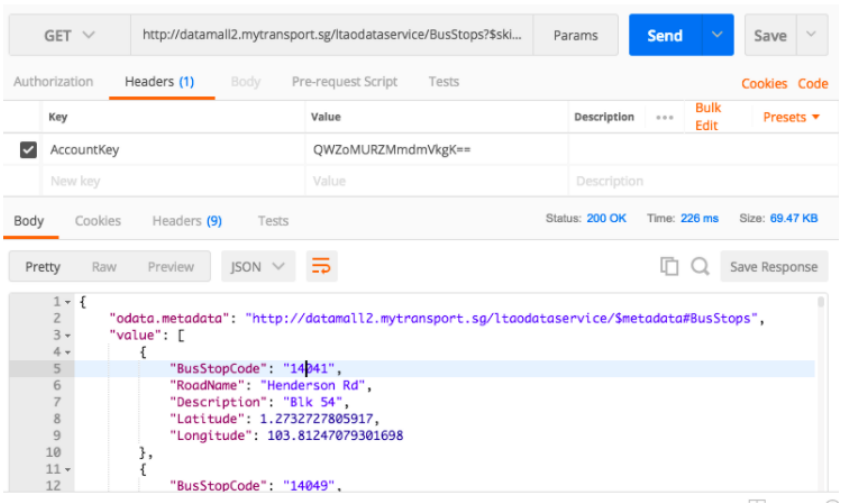
Let's pick a use case that's *a little more down to earth*: I live in Singapore and usually take the bus to the office. At times there's no free seats, sometimes not even a chance to get on the bus at all since it's packed with people standing. Singapore's Land Transport Authority (LTA) offers API access to data via its DataMall offering, with just a little coding we can **visualize** busy buses on Google Maps and **analyze** the peaks of a given day. In other words, we can transform something that looks just like the data snippets from the ISS example above -- into something like this:



We now have an interactive Google Map with all bus stops (within a 5 km radius from Raffles Place) that shows buses with standing-only capacity or the ones that are completely full. You can navigate through individual snapshots in 5-minute intervals with data from Friday, January 5th, 2018 here. And even more, we can compute a graph that illustrates how there's a lot more busy buses during peak hours - probably not a surprise:

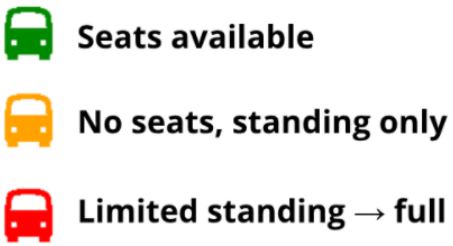


How did we get there? You'll find all code in my Github repository, but let's go through the steps here. After signing up at LTA's DataMall you'll receive an API Key which allows you to make API requests such as the following, where we're looking to get a list of all bus stops:



When going through the specs for the bus system I was excited to find out that besides bus stops, services, arrival times etc. there was a data field called '**Load**' that tells you

the current capacity of a bus. All we have to do is find a way to query the system frequently for all bus services to get the insight we're after. My choice of programming language for working with data is **Python**, so I've used the GeoPy library for identifying all bus stops that are within a 5 km radius of Raffles Place and PyCurl for making HTTP requests to query the API. The scripts produce data files with bus stop codes, locations and finally the current loads of the buses. I'm using **Shell-Script** wrappers to manipulate files and automate the process.

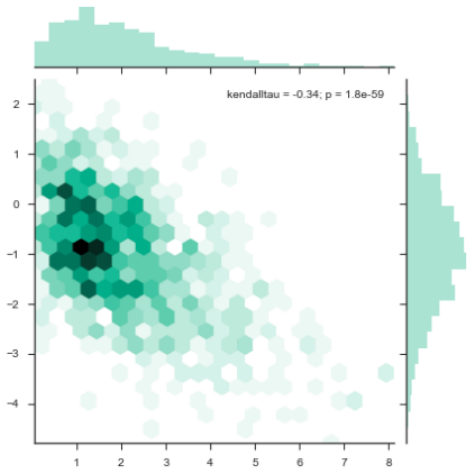


To create interactive Google Maps we need **JavaScript** for adding the custom markers with icons and text, and let's say we're not interested in the green buses. To automate this, I built this Python program that generates the JavaScript code for concatenation via Shell-Scripts:

```
{ position: new google.maps.LatLng(1.31389305598,103.881896111), type: 'busorange',
title: 'Aljunied Rd', text: ' Bus 137 (ETA 2018-01-07T11:34:40) standing only | ' },
{ position: new google.maps.LatLng(1.30069808015,103.886432997), type: 'busorange',
title: 'Mountbatten Rd', text: ' Bus 196 (ETA 2018-01-07T11:23:46) standing only | ' },
{ position: new google.maps.LatLng(1.30052638901,103.888313333), type: 'busorange',
title: 'Mountbatten Rd', text: ' Bus 196 (ETA 2018-01-07T11:23:00) standing only | ' },
{ position: new google.maps.LatLng(1.30314976548,103.884368453), type: 'busorange',
title: 'Mountbatten Rd', text: ' Bus 196 (ETA 2018-01-07T11:25:08) standing only | ' },
        ];

        // Create markers.
        features.forEach(function(feature) {
          var marker = new google.maps.Marker({
            position: feature.position,
            icon: icons[feature.type].icon,
            map: map,
            title: feature.title
```

The classic PyLab was an easy choice for plotting graphs, I've yet to figure out how Seaborn works, the examples look stunning! Needless to say, with the available data you can now do all sorts of analysis, compare week days to weekends, and even trending & predictive analytics. And ultimately, we might be able to improve service availability and customer experience - by using data!



What's next? Getting a taxi when it's raining is a challenge - both data sources, number of available taxis as well as current rainfall readings offer an API. Why don't you give it a try?

Report this

57 Likes

+47

Albert Sigit 6d
Sr. Business Group Manager Data Platform & BI
Are this using block chain?
Like Reply | 1 Reply

Uli Hitzel 6d
Executive Geek | Data Engineering & Automation
hi Albert! I've not used a blockchain service for this.
Like Reply |

Wen Jing Lim
Head in the IBM Cloud | API Economy and App Integration | ANZ

Indeed the LTA was far-sighted enough to capture data in all the buses, I would say it's a smart move for them when they left all the application creation to other companies and focused on making the data open. Thank you for making this sound so simple. I should try it and create an app for my usual commutes in a week.

Like Reply | 1 Like

Evgeni Makarov 3d ···
PhD in physics. Developing data analytics and software services.

not bad for Starbucks coding :)