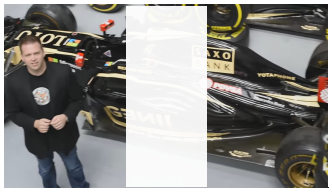
ADVERTISEMENT



Office 365 API - Hands-on
Microsoft Mechanics
See how Lotus F1 Team Developed Integrated dynamic app experiences

From where can I get data sets to practice modelling techniques?

dataset, model

sanket **Jan '15**

Hi all,

I am new to data analytics and am looking for data sets to work on for practicing modelling techniques.

Is there any source from where we can get the big data sets using which we can apply modelling techniques and sharpen modeling skills.

Regards,
Sanket

kunal **Jan '15**

Sanket,

Here are a few things which can help you get some practice datasets:

1. Look at some of the Kaggle knowledge competitions - Titanic, Bike sharing etc have decent sized databases for complete beginners.
2. You can look at the following article on Analytics Vidhya for a few other projects: **http://www.analyticsvidhya.com/blog/2014/11/data-science-projects-learn/**
3. There are a few huge datasets available on AWS for free **http://aws.amazon.com/public-data-sets/**
4. KDNuggets maintains a list of publocly available datasets as well: **http://www.kdnuggets.com/datasets/index.html**
5. Here is another repository from Google **http://www.google.com/publicdata/directory**
6. A similar question was asked on Quora some time back. Have a look at it here: **http://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public**

7. Here is a list from BigML **http://blog.bigml.com/2013/02/28/data-data-data-thousands-of-public-data-sources/**

Hope this gives you enough to practice on!

Regards,
Kunal

---

sanket **Jan '15**

@kunal Sir, Thanks 😄

Regards,
Sanket

---

vinisha **Jan '15**

Hi Sanket

There are a number of other websites that maintain data for analytical purposes. One of them is **http://stat-computing.org**

Apart from this website, every government department in the US maintains its data. So i suggest looking up at different government websites in the US to get free data sets.

another one is (though i have not used this yet)
http://www.google.com/publicdata/directory

You can also look up for the human genomics data..There is huge data available for free, or google books data or wikidata dumps etc etc.

Hope this helps

Regards

Vinisha

---

jon **Jul '15**

In addition to what has already been said:

- You can look at the **Kaggle** competitions. There is a competitive platform as well as some In class problems for academic use.
- You can also find some awesome data science projects for social impact on **DataKind**
- You can also look at open data applications available on **data.gov**

Hope this is useful

J

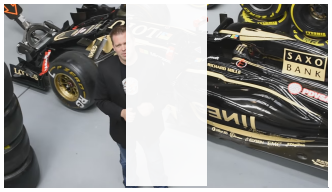
---

devanshrising **Jul '15**

Awesome Hackers!

kaggle is the good option as you get various approaches others have taken.

ADVERTISEMENT



Custom Office 365 Apps
Microsoft Mechanics
Lotus F1 Team builds a custom app w /Outlook integration for race crew

© Copyright 2015. Analytics Vidhya