dData
Our nowcasting challenge definition

Estimate all macroeconomic variables, for all countries, daily, using all available information

- **We aim to estimate all four tracks of the challenge** (production, consumption, trade and prices). By cross-cutting tracks we build holistic economy estimates using relations between indicators.
- **We assess all countries in Asia and the Pacific**, including those with extremely poor data as these are often the most vulnerable, and where nowcasting is most useful.
- **We nowcast variables daily** in order to integrate available data at its most granular level.
- **We create a scalable platform** so that additional information and new sources will continually improve our estimates.
The idea

To predict macroeconomic indicators by linking hundreds of diverse datasets and combining them to accurately nowcast economic measurements

Our technology links and aggregates economic datasets from a range of primary sources, from the World Bank’s Development Indicators to daily commodity price streams, helping users achieve faster, timely and more accurate insights into the world’s most pressing economic challenges

Our pilot test nowcast GDP, consumption, exports and inflation in the Philippines, Pakistan and Myanmar during the 2008 - 2009 financial crisis. We built a scalable solution, identified a community interested engaging with it, and are looking to implement our solution in late 2020 and early 2021.
The methodology

1. We convert each dataset into a common format that is comparable between sources. Where we have multiple sources of the same data point, we select the highest quality estimate.

2. We then use the available historic data to look at the relationships between variables. Not all data is available for each day and we predict missing variables daily, backfilling the missing data points.

3. We then identify what information we have at the current date and use the historic relationship of those variables with the variable of interest to nowcast the current economic measurement and predict future trends.
Vignette for nowcasting the financial crisis

We use data from the 2008 - 2009 financial crisis to assess the effectiveness of our predictions.

We assume knowledge of the sporadic data prior to 01/01/2008 and predict across the crisis for all the countries in the ADB dataset from that date, adding daily data as it would be available from that date.

We test to see if our model can accurately nowcast the macroeconomic shocks for the next 12-months and then predict the recovery until 01/01/2012.
Link data sources at different frequencies

To predict the crisis, we use ~150 macroeconomic variables at daily, monthly and annual frequencies.

**Daily stock exchange API**
- NASDAQ Global and Asia Indices
- Commodity markets (Agriculture & metals)

**Monthly indexes API**
- S&P index
- Financial rates including interest, earnings
- Trade (Freight and Price)
- Economist Big Mac Index

**Annual macroeconomic databases**
- World Bank Development Indicators
- WB Worldwide Governance Indicators
- WB Enterprise Surveys and Doing Business
- WEF Global Competitiveness Index
- IMF World Economic Outlook
- Economist Intelligence Unit
- Harvard Atlas of Economic Complexity
- Penn World Tables
- Quality of Government Dataset
The Philippines is an example of a country with high data availability:

- GDP and consumption data are available quarterly.
- Inflation and export data are available monthly.

Our backfill model finds a high degree of variation and tracks indicators effectively creating daily trend estimates from the low-frequency point data.

We predict at a quarterly time-frame using 24 variables, including shipping indices amongst other macroeconomic indicators.
The model accurately nowcasts near-term output

We find that the model accurately nowcasts the dramatic fall in exports during 2008 - 2009 as well as the first year of real GDP and consumption expenditure.

The model predicts some of the cost-push inflation spike but fails to assess its magnitude, rising only from 4% to 5% rather than the peak of 12%.

Overall the model nowcasts effectively and has some success in prediction except for inflation magnitudes.
Pakistan is an example of a county with medium data availability:
- GDP and consumption data are available yearly.
- Inflation and export data are available monthly.

Our backfill model finds a lower degree of variation and but still tracks indicators effectively creating a smooth trend out of infrequent GDP and consumption data.

In Pakistan only 14 variables are available on a quarterly basis to make predictions.
The model nowcasts well but fails to track variation

We find that the model has a low degree of variation in its nowcasting and prediction failing to rise and fall at the required magnitudes.

The nowcasting of the first 6 – 12 months appear accurate however, the model fails to predict the future acceleration in Pakistan growth, or exports and has difficulty with inflation magnitudes.

Overall the model is useful for nowcasting in medium data countries but not appropriate for prediction.
Myanmar is an example of a country with low data availability:

- GDP and consumption data are available yearly
- Export data are available monthly
- Inflation data is not available before 2008

Our backfill model finds a high degree of variation and performs well at tracking the scarce dataset including within year variation outside of the initial dataset.

In Myanmar only 8 variables are available on a quarterly basis to make predictions.
Identifies within year variation not in original dataset

The model's prediction identifies within year variation but fails to predict Myanmar's doubling of reported GDP between 2009-10, instead predicting an economic slump.

The nowcast of the first 6-months appear accurate and useful as data is unavailable, however, predictions past this point appear very different to the trend.

Overall the prediction accuracy appears much weaker but potentially helpful in nowcasting the first 6-months of missing macroeconomic data.
Assessment and near-term implementation plan

Our initial estimates are very rough and rely heavily on the high-frequency financial datasets, with more time and more computational capacity we would have linked more daily data and achieved a greater predictive accuracy.

We aim to improve on this initial vignette when modelling the COVID-19 crisis by:
1. Extending the model to this year, increasing the training dataset size from 8 to 20 years
2. Quality check existing values with alternative and subnational datasets
3. Automating the data formatting and nowcasting pipeline at all frequencies
4. Increase the number and heterogeneity of the data sources available

We plan to incentivize people to add data to the predictive model by creating a marketplace to buy and sell data in the common linked format. We are aiming to go-live with the dData exchange during late 2020 and early 2021.
The marketplace – https://atddata.com

We have created a peer-to-peer data sharing platform that crowdsources alternative datasets

1. The platform stitches together datasets and provides a single searchable output
2. Anyone can publish data and monetise its use through setting a price in Ocean tokens. We encourage a free-open, collaborative effort with a Wikipedia-like delivery model
3. Transactions quick and secure through online payment based on a e-token wallet system