Asia-Pacific
Economic Cooperation

Advancing Free Trade
for Asia-Pacific Prosperity

Project Summary Report: Capacity Building on Data Science Tools for Sector Regulators and Competition Authorities

APEC Economic Committee

July 2022

Asia-Pacific
Economic Cooperation

Project Summary Report: Capacity Building on Data Science Tools for Sector Regulators and Competition Authorities

APEC Economic Committee

July 2022

APEC Project: EC 05 2020A

Produced by
Philippines Project Team
Frontier Economics Pty Ltd
+61 3 9620 4488

For
Asia-Pacific Economic Cooperation Secretariat
35 Heng Mui Keng Terrace
Singapore 119616
Tel: (65) 68919 600
Fax: (65) 68919 690
Email: info@apec.org
Website: www.apec.org

© 2022 APEC Secretariat

APEC#222-EC-04.3

Table of Contents

Tables

Figures

Boxes

1 Introduction

1.1 Overview

Frontier Economics assisted the Asia-Pacific Economic Cooperation (APEC) and the Philippines Competition Commission (PCC) to develop a toolkit that provides a practical guide for sector regulators and competition authorities on the quantitative tools and techniques that they can use to gather and analyse data to inform their analyses (the Toolkit). The Toolkit was presented at workshop on 23 March 2022 (the Workshop).

The PCC's project team is composed of the PCC's Economics Office, which includes the following team members: Tea Jalin Ty, Maria Dominique Lucenario, Edgardo Manuel Jopson and Kirsten Dela Cruz.

This report provides an overview of the project, including developing the Toolkit, collecting and analysing primary and secondary data for the Toolkit, and conducting the Workshop. We also summarise the results from the evaluation surveys completed by Workshop participants.

1.2 Project objectives

As digital platforms continue to grow, and as traditional businesses make greater use of data driven insights, regulators will be increasingly called on to analyse big data as part of their regulatory decision-making process. It is imperative that regulators have the quantitative tools and techniques required to properly gather, clean and analyse complex data sets.

The objective of the project is to develop a Toolkit that provides a practical guide for competition authorities and regulators on the quantitative tools and techniques available to gather and analyse data in competition cases, studies and impact assessments. The Toolkit includes discussion of the strengths and limitations of the tools and techniques to assist competition authorities and regulators in deciding when particular tools and techniques should be applied.

The Toolkit will help regulators address the challenges they face in gathering and analysing increasingly large and complex datasets, and thereby help regulators to monitor and assess markets more effectively.

1.3 Structure of this report

The remainder of this report is structured as follows:

- Section 2 summaries the research activities that were conducted to develop the Toolkit
- Section 3 summaries the presentations and inputs provided at the Workshop; and
- Section 4 summaries survey feedback provided by Workshop participants.

2 Developing the Toolkit

This section provides a summary of the research activities conducted to develop the Toolkit.

2.1 Primary information

Primary data for developing the Toolkit was gathered through interviews with regulators from certain APEC member economies. The interviews covered how each regulator currently gathers and uses data to inform their competition assessments, and whether they had faced challenges with respect to their use of data.

The interviewees are set out in **Table 1** below.

Table 1: List of interviewees

Organisation	Name of interviewee	Date of interview
Philippines Competition Commission (PCC)	Tea Jalin Ty Maria Dominique Lucenario Edgardo Manuel Jopson	6 August 2021
Australian Competition and Consumer Commission (ACCC)	Peter Gray	12 August 2021
Competition and Consumer Commission of Singapore (CCCS)	Herbert Fung Weng Loong Kong	12 August 2021
Malaysian Competition Commission (MyCC)	Ismail Faruqi	17 August 2021
Competition Commission Hong Kong, China (HKCC)	Matthew Wong Eshien Chong	31 August 2021
Indonesian Competition Commission (KPPU)	Deswin Nur Enno Wiranti Inez Koerniawati	2 November 2021

Source: Frontier Economics.

Frontier Economics arranged and facilitated the interviews. Prior to each interview, we circulated a series of questions to the participants. These questions are set out in Box 1. As a result of the COVID19 pandemic, all interviews were conducted virtually.

Box 1: Interview questions provided to stakeholders

1. Do you use data in merger analysis? If so, where do you obtain the data? How do you analyse the data? Would you like more data? How would you like to be able to analyse extra data?
2. Do you use data in cartel analysis? If so, where do you obtain the data? How do you analyse the data? Would you like more data? How would you like to be able to analyse extra data?
3. Do you use data in analysis of abuse of dominance? If so, where do you obtain the data? How do you analyse the data? Would you like more data? How would you like to be able to analyse extra data?

Source: Frontier Economics.

Interviewees shared information on:

- the types of data they collect from market participants to inform their competition cases, studies and impact assessments
- the processes they use to collect this data, including data request templates, web scraping techniques, surveys, market studies, and by accessing data collected by other public bodies
- the data analytical tools that they had used or were developing, including models that can be used to detect cartel behaviour or similarity between documents
- the challenges they had faced in undertaking certain empirical analyses, such as using mapping software to examine the geographical extent of markets
- the processes they used to train staff in data analytics, including through their choice of institutional design (such as establishing a data analytics team) and through ongoing training.

A brief summary of key insights is provided in the table below.

Table 2: Key insights from interviewees

Interviewee	Summary of Key insights
Philippines Competition Commission (PCC)	• To inform their investigations, PCC gather data from market participants (including via merger notifications and data request templates), public resources (using web scraping techniques), market research databases, and other government bodies. Challenges arise when data is of poor quality, not sufficiently disaggregated, not in a usable format, or not digitised. • Complicated cases may involve complex statistical and econometric techniques, including difference-in-difference analyses, but most of the merger and cartel cases are resolved through relatively simply analysis of market shares and HHIs. • Identified the challenges they faced in undertaking certain empirical analysis, including using Geographic Information System (GIS) software to analyse spatial data for the purposes of information market definition analyses.
Australian Competition and Consumer Commission (ACCC)	• The ACCC established the Strategic Data Analytics Unit to help the development and deployment of advanced analytical techniques in competition investigations. Provided information on the specific analytical tools it had employed, including tools to detect cartel and bid rigging behaviour. • Outlined the integration of the Strategic Data Analytics Unit with the broader investigations team, and the challenges in communicating complicated statistical and econometric analysis to persons without the appropriate technical training.

Competition and Consumer Commission of Singapore (CCCS)	• The CCCS, in collaboration with GovTech, developed a cartel and bid rigging analysis tool. The tool has been successfully applied in a number of investigations. • To inform their investigations, CCCS gather data from market participants (including via merger notifications and data request templates), public resources (using web scraping techniques), market research databases, and other government bodies. Collection of data by public bodies in Singapore is comprehensive • The CCCS does not have a formal data analytics team. Some staff are more naturally inclined towards empirical work and will be used in cases that are likely to involve significant empirical work. In Singapore, there is a whole-of-government effort to educate and train government employees in data analytics, and CCCS staff participate in regular training sessions on data analytics techniques as part of this initiative.
Malaysian Competition Commission (MyCC)	• There is currently no merger control in Malaysia so the focus has been on detecting cartels and bid rigging. • To inform their investigations, MyCC gather data from market participants (including via merger notifications and data request templates), public resources (using web scraping techniques), market research databases, and other government bodies. • MyCC does not maintain a separate data analytics team. While it regularly employs standard competition econometrics, it does not typically employ complicated data science techniques as part of its investigations. • MyCC, in collaboration with Handshakes, developed a tool to detect bid rigging, primarily in the context of public procurement. The tool has been successfully applied in a number of investigations.

Competition Commission Hong Kong, China (HKCC)	• Hong Kong, China has only had a general competition law since 2015 and still has no general provision controlling anticompetitive mergers. Focus has been on cartels, anticompetitive agreements and abuse of market power • HKCC has used screening tools to review and assess large volumes of information to inform bid rigging assessments.
Indonesian Competition Commission (KPPU)	• Broad competition powers. Use of econometrics to examine competition issues though faces issues in accessing reliable data.

In addition to interviews, we also relied on our own experience in gathering and analysing data for projects that we have undertaken for public and private sector clients in the past

2.2 Secondary Information

Secondary data for developing the Toolkit was gathered from desktop research and including:

- Publications by international organisations – we referenced publications by the OECD (Organisation for Economic Co-operation and Development). For example, the OECD's "Review of Competition Law and Policy in Chile" (2010) and background paper titled "Big Data: Bringing Competition Policy to the Digital Era" (2016).
- Publications by regulators – we relied on reports published by the ACCC (Australian Competition and Consumer Commission) and CCCS (Competition and Consumer Commission of Singapore). The ACCC's report, titled "Digital Platforms Inquiry" (June 2019), is an inquiry on digital platforms e.g., Facebook and Google, and the CCCS's paper is titled "Data: Engine for Growth – Implications for Competition Law, Personal Data Protection, and Intellectual Property Rights" (2017).
- Academic papers – In carrying out our research, we covered a wide range of academic papers. Examples of papers covered include "Estimating demand for competition analysis – a statistical exploration, and some possible applications" (2020), "Data and market definition of Internet-based businesses" (2019) and "The role of quantitative analysis to delineate antitrust markets: An example. Blackstone / Acetex" (2005).

- Case decisions and literature – we reviewed papers outlining the decisions made by commissions in relevant and historical cases. For instance, we reviewed the Facebook and WhatsApp merger, Microsoft and LinkedIn merger and Google and DoubleClick merger cases.

2.3 Developing the worked examples

To support in the understanding and implementation of the techniques discussed in the Toolkit, we prepared a collection of datasets for practitioners to complete. The datasets contain hypothetical data that is similar to the type of data that may be collected by regulators over the course of an investigation. We outline the steps we took to build the datasets in this section.

Firstly, to construct the database, we started with nine petrol stations located in Melbourne, taking the latitude and longitude, address, and brand information. We then generated a set of potential customers by randomly drawing 300 locations near each of the petrol stations, applying a normal distribution to obtain the latitude and longitude.

To obtain the sample of customers for each store, we first determined which store each customer would choose. For each customer, we found the distance to each petrol station, excluded all stations 2.5km or further away from the customer, and then determined the relative distance by dividing by 2.5km. The stations are then randomly selected from the remaining stations, with the probability weighted by one minus the relative distance. For each store, we then randomly selected 100 customers who selected that store.

To obtain the retail data, we first constructed the prices charged for each store/product/week. We then created a market-wide price for each product for each week, allowing the prices of products R and S to be correlated. We then apply a small 'noise' term to the price set by each store. We adjust by the fuel index so that prices follow the trend observed historically. Finally, we make some adjustments: the price of Q at location A is set to a very high amount (i.e., withdrawn) from June 2020 onwards, and the price of Q at locations A and B are increased by 50% between June 2019 and August 2019.

For each week, for each customer, we determine how much of product P, Q and R they purchase, and at which stores. For each product the customer faces nine prices, one for each store, but they act on the basis of the effective price: multiplying by one plus the relative distance. In this way the effective price of a store 2.5km away from the customer would be twice as much as that of a store

located 0km away, even if the price charged were equal. The customer then chooses the store with the best effective price for the relevant product.

The customer now faces three effective prices – one for each product. Note that these may be from different stores. The customer then chooses quantities to maximise a utility function given the effective prices and the budget constraint. The preferences are Cobb-Douglas over Q and a composite good: CES preferences over R and S. In this way, customers exhibit imperfect substitution between R and S, and spend a constant fraction of income on Q.

We repeat this for each week and customer, and then aggregate to construct the retail dataset.

We then create 'dirty' datasets by making several errors. For example, we randomly duplicated observations, changed some product names to lower case letters, allowed store H to use the 'week starting' rather than 'week ending' convention for dates, and changed the latitude or longitude of stores.

3 Overview of workshop

3.1 Workshop Introduction

On the 23rd of March 2022, the PCC organised an online Workshop commissioned by APEC titled Capacity Building Workshop on Data Science Tools for Sector Regulators and Competition Authorities amidst the New Normal. The Workshop was held on Microsoft Teams from 9am to 11.30am (GMT+8).

For the workshop proper, the project team reached out to four panel speakers from different jurisdictions to present and discuss their experience in analyzing big data and their respective tools for analysis. There were four panel speakers representing CCCS, ACCC, the Department of Trade and Industry of the Philippines, and the MyCC. The workshop was well-attended by 268 participants from the original target of 60 on-site attendees. The participants came from 11 APEC member economies including the Philippines, Brunei Darussalam, Japan, Mexico, the United States, Canada, New Zealand, Malaysia, Singapore, Australia, and Chinese Taipei.

Gender-disaggregated data among workshop participants were also collected. The workshop targeted gender parity among invited workshop resource persons, and also ensured that there are no gender restrictions regarding suppliers, partners, or participants. In this endeavour, the workshop was able to achieve its gender targets, with 125 participants or 47 percent of its attendance consisting of women from various APEC member economies, and by having one female expert in a panel of four speakers (25 percent).

Meanwhile, some challenges encountered during the conduct of the international workshop include reaching the target of having representatives from 15 APEC member economies to attend the event. While the virtual workshop was well-attended, the participants only came from 11 different APEC member economies as this may be attributed to several factors such as the time zone difference and possible decline in interest due to the limitations of conducting the Workshop through an online platform.

As a result also of COVID-19, the Project Team was required to organise a virtual Workshop instead of a physical one. This led to occasional technical difficulties experienced by the speakers. However, the IT team was on standby and quick to rectify any technical issues that arose. In addition, a virtual dry run involving the PCC, the IT team, and presenters was held in the days preceding the Workshop

to ensure that any major technical issues were dealt with prior to the main event. Going forward, the project team will continue to incorporate dry runs in project plans and ensure that a skilled IT team is on standby during the event.

The workshop program including the project briefer were provided in the Annex. A six-month post-APEC Data Science Workshop evaluation instrument was also appended to document the participants' assessment in terms of increased awareness and appreciation in data science tools, new knowledge gained, increased interest and engagement in data science tools, and level of applicability to their regulatory functions.

3.2 Opening Remarks

The Workshop began with Dr. Arsenio M. Balisacan, the Chairperson of the PCC, delivering opening remarks regarding the importance of effective data science tools and techniques for regulators and competition authorities to be able to leverage when carrying out their regulatory functions. The increasing dominance of these digital platforms has the potential to result in market players abusing their dominance and exploiting consumers. Therefore, sector regulators and competition authorities must be adequately equipped with the necessary data science tools to be able to analyse and convert data into useful information with valuable insights that will ultimately feed into the decisions made by policy makers. This Workshop aims to facilitate the uptake of data tools and techniques amongst regulators and competition authorities.

3.3 Presentation Highlights

3.3.1 Overview of the Data Science Toolkit

Data Science for Competition Policy Toolkit - Consultants from Frontier Economics, who advised APEC on developing a Data Science Toolkit for Competition Assessments, then provided an overview of the Data Science for Competition Policy Toolkit, its contents and potential application for sector regulators and competition authorities. Bob Bartels and Ehson Shirazi discussed the different types of data that are generated from digital platforms. Mr Shirazi explained that whilst this surge in big data delivers benefits for consumers and businesses e.g. by improving the quality of goods for customers, encouraging innovation and consumer choice, improving supply chain efficiency for businesses, industry regulators and competition authorities need to have the necessary knowledge and Tools to be able to assess and monitor these markets effectively. Prof

Bartels continued by providing the audience with a summary of the contents of the Toolkit. He provided examples on the types of techniques and tools that can be applied in the process of gathering data, cleaning data and analysing data in the context of regulators and competition authorities performing their regulatory functions. Examples of the data science tools are: market modelling, regression analysis and utilising the Geographical Information System (GIS) software.

Key Insights

- **Big data platforms have changed the landscape of competition**. Platforms such as Google and social media networks have made it easier to create and extract data. With more people and businesses connecting to the internet competition authorities must tap on this wealth and source of information previously unimaginable.
- **Big data benefits businesses and consumers alike**. The use and analysis of big data allow businesses to innovate, adopt to rapidly changing conditions, and streamline processes and improve efficiency. Ultimately, these lead to better products and services for consumers.
- **Regulatory agencies must step up to the challenge of big data**. Big data analytics can aid in monitoring anti-competitive behaviours and abuses of dominant position. Regulators may strengthen its enforcement and come up with more informed decisions using data analytics and big data.

3.3.2 Panel discussion highlights

The panel discussion session which followed involved speakers from competition authorities and regulators in the APEC region. The panel consisted of representatives / speakers from competition authorities and regulators from Malaysia, Australia, Singapore, and Philippines who shared their practical experience using data science tools in the context of carrying out their regulatory functions. Dr Philip Williams from Frontier Economics moderated the panel discussions as well as the Q&A session that followed.

Enhancing the comparison of documents in enforcement - The discussion commenced with Kong Weng Loong, the Deputy Director of the Competition and Consumer Commission of Singapore (CCCS), sharing the CCCS's experience in using a data science tool for enhancing the process of comparing documents e.g. tender proposals. The 'text similarity' tool discussed uses text analytics to compare the similarities between documents at both a sentence and document

level. It is used to enhance enforcement capabilities in analysing documentary evidence and detecting suspiciously similar documents e.g. detecting similarities in tender proposals of competitors. Overall, this Tool supports in the process of detecting suspiciously similar documents and sentences and complements quantitative analysis e.g. analysis of bid data to detect bid-rigging and extracted price information.

Using natural language processing to extract value from consumer complaints - Next, Kristen Osborne, a Senior Analyst at the Strategic Data Analysis Unit at the Australian Competition and Consumer Commission (ACCC), discussed applying Natural Language Processing (NLP) to extract value from consumer complaints at the ACCC. Instead of manually reviewing complaints in the form of free text at the ACCC, the team rely on NLP to feed information to a model to teach it to understand what the ACCC is seeking. An 'off-the-shelf' model was built using Spacy, a natural language processing tool within Python, but was subsequently customised. Ms Osborne concluded the presentation by explaining the benefits of NLP-based tools, such as its ability to automate and improve the analysis of free text and that this process can be achieved using free, open-source software. However, she caveated this by saying that similar to any Tool, it requires being designed and used by individuals who possess the skills to be able to do so.

Philippine National Standard: Guidelines for electronic commerce transaction - Edgardo D. Del Rosario, a trade industry development specialist at the Bureau of Philippine Standards Philippine Department of Trade and Industry (DTI), then discussed the technical references on online / electric commerce transactions. He identified that there has been a significant increase in the number of consumer complaints related to online transactions since the outset of Covid-19. As a result, the DTI published a technical reference on guidelines for electronic commerce transactions in the Philippines (PNS 2155:2020). The guideline can be used by e-market places, 3rd party services supporting the operations of online transactions and e-retailers. The document enables online marketplaces to operate in a more customer-friendly manner and ensures that customer-centric business policies and processes are put in place to increase customer satisfaction. This document is a guideline on the key activities surrounding online transactions e.g. pre-purchase, during, and post-purchase activities.

Leveraging data science tools for the indication of bid-rigging cartels - The final panel speaker was Ismail Faruqi Abdullah, who is the acting head of business and economic division at the Malaysia Competition Commission (MyCC). He discussed the data science tools that the MyCC uses to identify bid-rigging cartels. For instance, MyCC collaborated with Handshakes to develop a

tailor-made data science tool to investigate bid-rigging cartels. Using the tool, MyCC can identify and extract company information including details on shareholders, map out the relationships between companies or individuals to uncover potential conflicts of interests and expedite the investigation process, which is both cost and time saving.

Key Insights during the Panel Discussion - Regulators and competition authorities from Singapore, Australia, Malaysia, and the Philippines share experiences on the use of data science for competition regulation.

- **Building a data analytics and management team is a must.** Competition authorities must assemble a competent data analytics team which could complement the agency's functions and add value to data for end-users such as regulators and consumers.

- **Continuously develop and exploit a host of data analytics tools available for regulators.** There is a wealth of resources and techniques that can be accessed and further developed by regulators such as text similarity detectors and natural language processing. The use of such tools can streamline detection of anti-competitive actions, lead to more accurate assessments, and save a lot of resources for the agency.

- **Establish guidelines on data collection, security, and protection.** Part of the responsibilities of regulators is the protection of consumers' sensitive and personal data. Regulatory agencies must spearhead creating and advocating for rules and standards for the handling of data. These standards may also be useful for agencies to ensure that data collection processes are efficient and could address the needs of authorities.

- **A culture of data literacy in regulatory agencies is a must.** While specialization in data analytics from top to bottom is not expected, competition authorities must at least aim to be data-literate across the board, with all members possessing an understanding of data and its varied use. Regulatory agencies must work to build a capable staff and tap on a pool of human resources that is proficient in data analytics.

3.4 Open forum / Q&A highlights

Following the panel discussion, Prof Philip Williams moderated the Q&A session. Examples of the types of questions raised include "what are the different types of programming software used to develop text analytics tools?", to which Mr Kong elaborated on the use of Python at the CCCS in text analytics. In addition, Ms Osborne was asked about her experience dealing with colleagues who are less proficient in natural language processing and other data science tools. All of the speakers were asked questions about their presentations.

3.5 Closing remarks

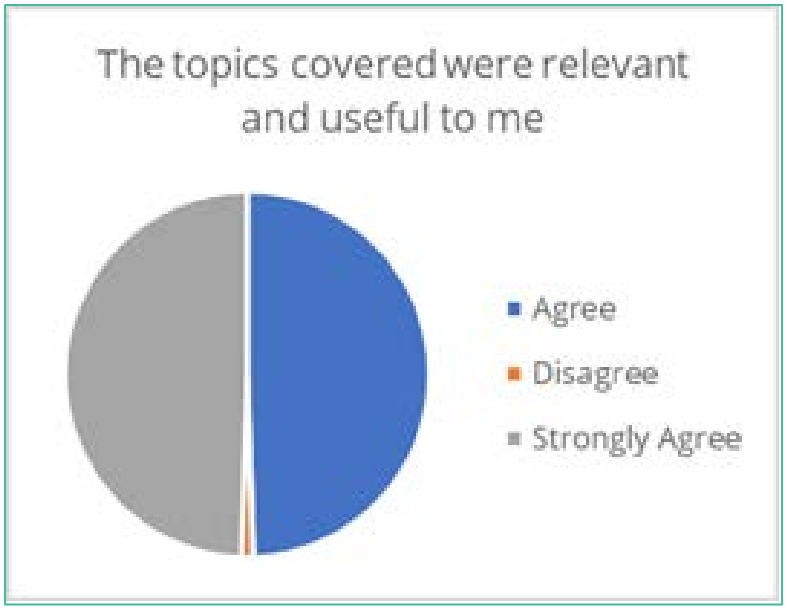
Lastly, Kenneth V. Tanate, Executive Director of the PCC, delivered his closing remarks. On behalf of the PCC, Mr Tanate thanked all speakers and attendees, and emphasised the importance of this event in supporting competition authorities and regulators in carrying out their regulatory duties effectively.

4 Evaluation form results

The PCC released an online evaluation form for attendees to provide their feedback on the APEC Data Science Workshop. A total of 190 responses were submitted out of the 268 registered participants from 11 APEC member economies, which include the Philippines, Brunei Darussalam, Japan, Mexico, the United States, Canada, New Zealand, Malaysia, Singapore, Australia, and Chinese Taipei.

99% of participants found that the Workshop covered topics that were relevant and useful to them. One individual wrote "I thought there was a good combination of new big data examples, like the textual analysis, and basic tools provided in the Toolkit." Another said, "Amazing panel, presentations, and topics for discussion. I took a lot away and found them very relevant."

Figure 1: Results of survey statement "the topics covered were relevant and useful to me"



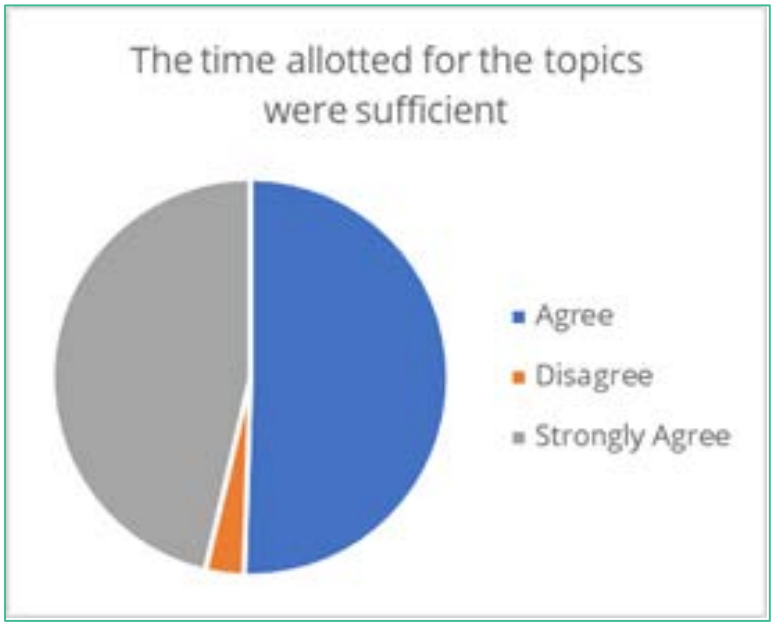
Source: Frontier Economics analysis

The form also provided participants with the opportunity to present feedback on each panellists' individual presentations.

100% of individuals who attended Frontier Economics' presentation on Data Science for Competition Toolkit agreed that Prof Bob Bartels and Mr Ehson Shirazi had clearly discussed the Toolkit. Most responses stated that both speakers were concise, knowledgeable, and well-versed with the Toolkit.

Regarding the presentations delivered by the panellists, 100% of participants agreed that Mr Kong Weng Loong (Singapore- CCCS) and Ms Kristen Osborne (Australia- ACCC) had clearly discussed their topics "Enhancing the Comparison of Documents in Enforcement" and "Using Natural Language Processing to Extract Value from Consumer", respectively. Mr Kong and Ms Osborne received specific feedback saying that their presentations were "excellent", "clear and informative" and "included good examples". In addition, 99% of attendees found that Mr Ismail Faruqi Abdullah clearly discussed the topic "Data Science Tools for the Indication of Bid-rigging Cartels". He received feedback saying that his presentation was "very clear and informative" and more specifically, that he had "expertly demonstrated the tools they use in their economy in order to indicate and investigate bids for any potential conflicts." Finally, 98% of attendees agreed that Mr Edgardo D Del Rosario had clearly discussed the topic "Philippine National Standard: Guidelines for Electronic Commerce". Overall, he received feedback saying that he had explained the topic well but also received suggestions saying that he should have simplified the topic as it was "too technical" in some instances.

Figure 2: Results of survey statement "the time allotted for the topics were sufficient"



Source: Frontier Economics analysis

Whilst 97% of respondents agreed to the statement that "the time allotted for the topics were sufficient", numerous individuals suggested to "allocate more time for presenters to expand on their topics". Ultimately, 99% of respondents agreed that the time allotted for the Q&A / Open

form was sufficient. Furthermore, in response to the question “How would you rate the platform used (MS Teams)?”, 87% of respondents rated MS Teams as either “very good” or “excellent”.

5 ANNEX

5.1 Workshop Project Overview and Agenda

Capacity Building Workshop on Data Science Tools for Sector Regulators and Competition Authorities amidst the New Normal

23 March 2022 | 09:00am -12:00nn | Microsoft Teams

PROJECT OVERVIEW

Our growing reliance on digital platforms amidst the COVID-19 pandemic has made a significant impact on global economic activity, especially for micro, small and medium enterprises (MSMEs). These digital platforms had to accommodate the increase in demand for safe, dependable, and affordable products and services, as consumers adopted to physical distancing measures aimed to prevent the spread of the virus. Through these platforms, big businesses and some MSMEs were able to adjust and continue operations despite the uncertainty and risk arising from the challenges brought about by this New Normal.

In the pursuit of these digital platforms to support changing consumer behavior, large sets of data or 'Big Data' are collected and processed, allowing automatic recalibration of prices, targeted advertising and promotions, efficient web search, among other technological developments. With these technological advancements, sector regulators and competition authorities may not be equipped with the necessary tools to effectively monitor and assess firm behavior despite the availability of Big Data. This can be seen among Asia Pacific Economic Cooperation (APEC) member economies that structurally operate with limited resources, technical machinery, and experience. Given the increasing number of firms that are quickly adjusting to the digital landscape, there is a need for regulators to develop more robust and systematic methods for gathering and analyzing data. With this, regulators will greatly benefit from equipping themselves with cutting-edge analytical tools and techniques in assessing the growing dominance of digital market platforms, and whether there are significant entry barriers that impede competition in various markets affected by this new digital environment.

The APEC Secretariat and the Philippine Competition Commission (PCC), through Frontier Economics, presents in this online workshop its project entitled 'Toolkit on using Data Science Tools for Competition Assessments'. This toolkit aims to enhance the appreciation of the link between the availability of Big Data and data science tools for their respective regulatory functions and to serve as a detailed guide on the appropriate use of various techniques for quantitative analysis in the field of competition policy.

Capacity Building Workshop on Data Science Tools for Sector Regulators and Competition Authorities amidst the New Normal

23 March 2022 | 09:00am -12:00nn | Microsoft Teams

WORKSHOP AGENDA

09:00 – 09:05	**Welcome and registration process**
09:05– 09:10	**Opening Remarks** *Arsenio M. Balisacan* *Chairperson, Philippine Competition Commission*
09:10 – 09:15	Virtual Photo Opportunity
09:15 – 09:45	**Data Science for Competition Policy Toolkit** Overview of Data Science for Competition Policy Toolkit, its contents and application for sector regulators and competition authorities *Bob Bartels, Frontier Economics* *Ehson Shirazi, Frontier Economics*
9:45 – 10:45	**Panel Discussion** Sharing practical experience in the use of data science tools in relation to the various activities of competition authorities and sector regulators **Enhancing the comparison of documents in enforcement** *Kong Weng Loong, Deputy Director,* *Competition and Consumer Commission of Singapore* **Using natural language processing to extract value from consumer complaints** *Kristen Osborne, Senior Analyst, Strategic Data Analysis Unit, Australian Competition & Consumer Commission* **Philippine National Standard: Guidelines for electronic commerce transaction** *Edgardo D. Del Rosario, Trade Industry Development Specialist, Bureau of Philippine Standards, Philippine Department of Trade and Industry* **Leveraging data science tools for the indication of bid-rigging cartels** *Ismail Faruqi Abdullah, Acting Head, Business and Economics Division, Malaysia Competition Commission* *Moderator* *Philip Williams, Frontier Economics*
10:45 – 11:25	**Open forum**
11:25 – 11:30	**Closing Remarks** *Kenneth V. Tanate* *Executive Director, Philippine Competition Commission*

Master of Ceremonies
Maria Dominique A. Lucenario, PCC

5.2 Project Briefer

Digital platforms have been proliferating around the world. While it has been established that social media and web search giants such as Facebook, Twitter, and Google have shaped our lives in the turn of the century, consumers and businesses in the past decade have also witnessed the rise of other digital platforms such as online shopping platforms, app-based transportation and delivery services, e-wallets, and cryptocurrency. This growth has been driven by technological innovations that have increased online connectivity, improved the availability of supporting infrastructure such as cloud computing and data storage, and expanded data processing capabilities. Digital platforms have leveraged network effects to grow their customer base and expand their service offering. More recently, the COVID-19 pandemic has also contributed to the growth of these digital platforms, with many businesses and customers moving towards online transactions to comply with physical distancing measures introduced by Governments.

The growth in digital platforms has seen a commensurate increase in the data generated by these platforms. Platforms collect large amounts of data from platform users. This may include, for example:

- *Personal data* – this is data required to confirm the identity of users on the platform and is typically collected during the signup process, such as name, phone number, address, etc.
- *Operational data* – this is data which the platform requires to provide its matching and connecting service, such as product listings, service offerings and financial information including credit cards and e-payment details
- *Search data* – this is data on what products and services customers are searching for on the platform, and the location and timing of the searches
- *Transactional data* – this is data on what products and services customers have purchased on the platform, when they were purchased, and how much was paid by the customers.

Regulators and competition authorities must ensure that they have the necessary knowledge and tools to assess and monitor markets effectively. Anti-competitive practices in the market, which include anti- competitive agreements, abuse of dominant position, anti-competitive mergers and acquisitions (M&As), and government regulations and policies that impede markets from competing effectively for better prices and product quality, have increased in complexity as

technology continues to advance. Regulators who fail to adapt to these changes will cause a substantial decrease in competition for certain markets. With the rapid growth in digital platforms, there is a need to develop more robust and systematic methods for gathering and analysing large and complex datasets. The increase in the number of firms who are rapidly adjusting to the digital economy only emphasises the urgency for regulators to keep up with these developments, especially since greater cross-border transactions are occurring within the Asia Pacific Economic Cooperation (APEC) region.

This Toolkit provides a practical guide for competition authorities and regulators on the quantitative tools and techniques available to gather, clean and analyze data in competition cases, studies and impact assessments. Within the data gathering section, the Toolkit considers the tools and techniques that can be used to gather data through merger notification templates, web scrapping techniques, accessing government information, market studies and surveys. The data cleaning section then explores cleaning up data that may contain errors, inconsistencies, or missing values, that regulators and competition authorities may need to use for empirical analysis. Finally, the data analysis section considers the empirical techniques that can be used to analyze typical competition issues encountered by competition authorities and regulators. The issues considered include identifying collusion, defining markets, assessing market power, predicting outcomes of horizontal mergers, and determining appropriate penalties for anticompetitive conduct.

The Toolkit also includes discussions of the strengths and limitations of the various tools and techniques to assist competition authorities and regulators in deciding when particular tools and techniques should be applied.

Examples of the empirical methods from the Toolkit include:

1. Using regression analysis to detect cartels

   Pricing data can be used to determine when collusion has succeeded in raising prices above the levels that would otherwise have been obtained. The analysis should compare prices in the market that were determined outside the period of the alleged cartel conduct with prices that were determined during the period of the alleged conduct, while controlling for changes in prices that are related to changes in costs over time. If prices determined during the period of the alleged cartel were found to differ systematically from prices determined outside this

period, this may support a finding that the alleged cartel caused prices to be different from what they otherwise would have been.

Undertaking this analysis involves 4 key steps:

a) Specifying the regression model

b) Gathering and cleaning the data

c) Estimating the parameter coefficients and testing robustness

d) Interpreting the regression results

2. Using mapping software to understand the geographical extent of a market

Competition authorities are concerned with the exercise of market power. To assess market power, it is often helpful to define the relevant market. The Toolkit provides a guide on the use of geographical information system (GIS) software in the context of market definition, in particular the geographical extent of a market participant's influence on other market participants. GIS is a powerful tool that can help competition authorities understand local markets in situations where markets are likely to be strong local elements, for example supermarkets or petrol retailers.

There are many calculations that may be performed using GIS that are otherwise impractical, for example calculating the share of a suburb within 5 kilometres of a retail store location.

The use of mapping software (GIS) involves several facets:

a) The representation of locations and areas in a spatial environment

b) Performing calculations using the spatial objects

c) Presenting results spatially

3. Modelling different outcomes to analyze the impact of implementing new policies

The Toolkit describes how regulators can utilize market modelling to project different market outcomes in different settings or scenarios. Market modelling can be used to assess the impact of new policies, M&As, or even changes in fuel prices in various markets.

For instance, Frontier Economics possesses an electricity investment model, called the WHIRLYGIG, that regulators can use to assess the long-term impacts of changes in the

electricity market. Examples of these would include the deployment of new generators and interconnectors added to the energy system each year, or the introduction of electricity policies (e.g. renewable energy targets and carbon emissions constraints) which impact market participants. The model determines the efficient (least-cost) operation and investment in an electricity market over a long-term investment horizon. Regulators will be able to determine the optimal size, type, location, and timing of new generation capacity and/or network changes to meet demand over the analysis horizon.

The different tools and techniques set out in this Toolkit are not necessarily new, however its application to more complex data is what the Toolkit offers. They have been (and continue to be) used by competition authorities and regulators around the world. The theory underpinning the analysis of data is grounded in established principles of economic theory, econometrics, and statistical analysis. While many of the tools and techniques that are discussed in the Toolkit can be applied across different sectors and industries, the Toolkit includes specific discussion and examples covering their application in the context of digital platforms.

5.3 Six-month post-APEC Data Science Workshop Evaluation Form

Demographic questions and other personal information:

Questions	Answers
Name	
Age	
Sex	
Organisation	
Designation	
Economy	
Email address	

Post-APEC Data Science Workshop Evaluation Questions:

Questions	Answers
1. Do you feel that the Workshop increased your **awareness** of the types of Data Science Tools available for you to perform your regulatory functions?	☐ Yes ☐ No
2. Has the Workshop enabled you to have a better **understanding** of the Data Science Tools discussed by the panellists at the Workshop?	☐ Yes ☐ No
3. Has the Workshop enabled you to have a better **appreciation** for the Data Science Tools discussed by the panellists at the Workshop?	☐ Yes ☐ No
4. Since the Workshop, are you **better able to engage** in discussions regarding the Data Science Tools shared by panellists at the Workshop?	☐ Yes ☐ No
5. Has the Workshop prompted you to **initiate the use** of any of the various Data Science Tools (discussed at the Workshop) in undertaking your regulatory functions?	☐ Yes ☐ No
6. Since the Workshop, have you applied any of the **empirical methods** discussed in APEC's Data Science Toolkit?	☐ Yes ☐ No
7. Since the Workshop, have you relied on any **new Data Science Tools** to carry out your regulatory functions?	☐ Yes ☐ No
8. Since the Workshop, have your **data gathering** processes improved?	☐ Yes ☐ No
9. Since the Workshop, have your **data cleaning** processes improved?	☐ Yes ☐ No

10. Since the Workshop, have your **data analysis** processes improved?	☐ Yes ☐ No
11. Since the Workshop, have you applied any **insights and international best practices** on the use of data science tools as discussed by the speakers?	☐ Yes ☐ No
12. Since the Workshop, has there been an **increased interest in international collaboration,** specifically on the use of data science tools, at your agency?	☐ Yes ☐ No
If you have any further comments, please provide them in the box below.	