

Data-lympics 2019

28-29 January 2019



Agenda

- Welcome remarks
- Rundown
- Rules
- Judging
- Challenges
- Q&A

Welcome remarks

Jennifer Ho

Leader of China/Hong Kong
Digital Risk Solutions Services



Rundown



Rundown

Day 1: 28 January 2019 (Monday)

| | |
|-------|----------------------------|
| 9:00 | Registration |
| 9:30 | Welcome remarks |
| 10:00 | Competition briefing |
| 11:00 | Competition starts |
| 13:00 | Group photo taking |
| 13:15 | Self-served lunch ready |
| 18:00 | End of Day 1; Venue closed |

Day 2: 29 January 2019 (Tuesday)

| | |
|-------|--|
| 9:00 | Registration |
| 9:30 | Competition continues |
| 12:00 | Solution submission |
| 14:00 | Announcement of shortlisted teams |
| | Please note that all teams should come back to the venue starting 13:30 |
| 15:00 | Presentation |
| 17:15 | Awards announcement |
| 17:30 | Closing remarks |

Rules



Rules



The PwC's Data-lympics is a 2-day event that allows participants from universities to have an opportunity to demonstrate how data & analytics can be used to address business challenges. In addition to the given datasets, teams are free to use any open data or crawl data from the Internet.



Participants **MUST** bring and use their own personal computer devices. Each team should work on exactly one challenge selected from the list. For this Data-lympics, teams cannot propose their own topics. All teams must begin coding at the same time. All development work must be done within the development period stated in the event schedule.



PwC will own the rights to the champion's projects (i.e. source codes and other deliverables) the participants create during the Data-lympics. The copyright, intellectual property and other rights of all datasets as provided in the Data-lympics are solely owned by the PwC. Participants are required to sign a non-disclosure agreement for the data provided during the Data-lympics.



8 teams will be selected based on technical aspect of their solution to perform final presentation. The final presentation will be of **7 minutes** with **3 minutes** question and answer.



The organising committee reserves the rights to disqualify the participants for late, absent, or improper dressing and behavior during the event.

Awards

The top 3 winning teams will be awarded with two-months internship or one-year placement programme*, plus:

| | |
|-------------------------|------------------------|
| Champion | HK\$20,000 cash reward |
| First runner up | HK\$12,000 cash reward |
| Second runner up | HK\$5,000 cash reward |

** With PwC Hong Kong and Mainland China's Risk Assurance's Data & Analytics team*

Besides the above rewards, we will also be selecting team for following awards:

- **PwC Data Hunter Award**
- **Most Innovative Award**
- **Best Presentation Award**

Housekeeping notes

All participants will be required to sign an agreement in relation to non-disclosure, intellectual property rights of the submissions and disclaimers.

A T-shirt will be provided to each contestant for identification purpose when you sign in. You are most encouraged to remain dressed in the T-shirt provided throughout the event.

Please turn your mobile phone or any devices to silent mode, and keep your noise level down during the competition in consideration of other contestants.

Food and beverages will be provided to you throughout the day at the venue. Alternatively, you are free to make your own meal plans (Food items with any forms of "soup" is not allowed in the competition room – please consume them at the public common area).

Please take care of your personal belongings.

Judging



Judging

Judging team



Jennifer Ho
Partner



Kristine Chung
Partner



Thomas Fu
Director



Chris Mo
Senior Manager



Jason Tung
Senior Manager



Patrick Mulholland
Senior Manager



Jordan Mootoosamy
Manager



Henry Yau
Manager



Camellia Zhang
Manager

Evaluation criteria for top 8 shortlisted teams will be based on technical accuracy of their solutions.

The following criteria will be used in evaluating their final presentations:

| Evaluation criteria for presentation | Points |
|--|--------|
| Relevance <ul style="list-style-type: none"> Is the solution relevant to the business issue? Is the solution relevant to Hong Kong/China? | 30 |
| Impact <ul style="list-style-type: none"> Is the solution adding value to the business? | 10 |
| Innovation <ul style="list-style-type: none"> Does the solution introduce new ideas or methods? Does it make use of open data? | 20 |
| Technical achievement <ul style="list-style-type: none"> Readiness to use (Technical accuracy e.g. F Score) Ease of use | 30 |
| Communication and presentation | 10 |

Challenges



Challenge 1: Computer vision challenge



Background and the need of the challenge

Many industries these days are having a desire to understand the activities and interests of consumers for statistical purpose adds value to develop marketing strategies and decide resource allocation.

By analysing a video clip, we would like you to develop a real time object detection algorithm which will be able to capture and recognise car registration number. The algorithm should be able to transform car registration number data into digitalised format and import either in database or export file as well as to be able to plot a graph to reflect live situation.



Data provided

We will provide files in a website once your team have signed the non-disclosure agreement:

1. Filename Hong Kong Car Number Plates – which is a list of sample object images for training purpose.
2. Filename TrainingVideo (HD_video) – which is a sample training video clip as well as for the evaluate your result.



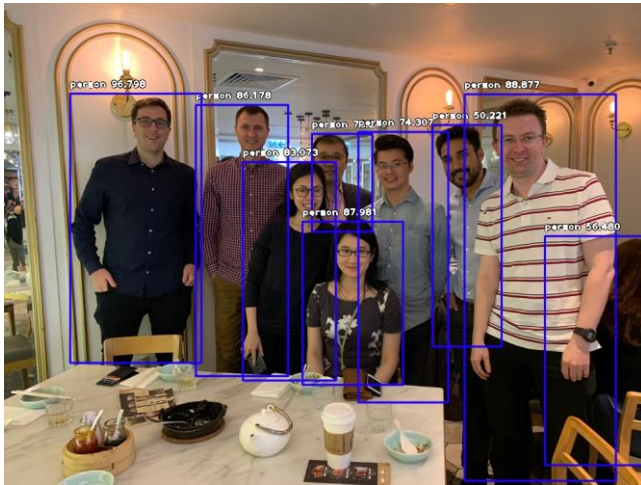
Submission

You will need to submit the following to **data-lympic@hk.pwc.com** by 12pm of 29 January 2019:

1. Source code and a list of library used by your solution
2. The expected output result of your solution
3. Operation guide

Challenge 1: Computer vision challenge

Object detection



Required to recognise each of object and generate statistic records through real time object detection



Required to recognise each of object and convert pictures or text through real time object detection



5412 7512 3456 7890

Challenge 1: Computer vision challenge

Submission checklist

| | Materials required | Details |
|---|--|---|
| 1 | Source code and dependencies | <ul style="list-style-type: none">• All prerequisite associated software(s)• The database being used and its settings• All associated library / modelling / algorithm used in coding• The algorithm MUST NOT only can train 1 and only 1 object• The algorithm MUST allow to support dynamic input source(s)• The algorithm MUST has an mechanism for detecting specific object(s) which can be supported either user input or a list of targeted objects for finding particular number(s).• The algorithm SHOULD has an mechanism for sending notification |
| 2 | Operational guideline | <ul style="list-style-type: none">• All necessary input and output parameters of the algorithm• All limitations of the algorithm and any potential problem(s)• Step-by-step operational procedures (End-to-end data flow diagram and procedures)• Procedure of the embedment of the algorithm into the camera (e.g. Dashcam, GoPro, Papago!) (Optional) |
| 3 | CSV or XLSX file containing details of detected objects | <ul style="list-style-type: none">• Total number of detected objects during the video period in Numeric format• Each of the recorded content - targeted car plate number in TEXT format• Sending notifications to a specific email when detecting the targeted object• Capturing timestamps whenever an ACTION occurred <p>Below is an illustrated output:</p> <pre>Total no. of car plates detected: 5 AB1234, 12:30:01 BC2345, 12:32:10 DE3456, 12:38:12 FG4567, 12:43:29 GH6789, 12:53:40 </pre> |

Challenge 1: Computer vision challenge

Round 1 evaluation: Technical accuracy

| Evaluation criteria | Formula | Weighting |
|--|--|-----------|
| Compare the accuracy of the total number of detected objects within the video period | $100\% - \frac{\text{Abs (Actual answer – Submitted no.of detected objects)}}{\text{Actual answer of detected objects}}$ | 40 |
| Compare detected car registration No. accuracy | $1 - \frac{\text{No.of incorrect car plates}}{\text{No.of correct car plates}}$ | 40 |
| Passing rate would be 75% or above | NA | 20 |

Challenge 1: Computer vision challenge

Round 2 evaluation: Presentation

| Evaluation criteria | Details | Points |
|---|---|--------|
| Relevance <ul style="list-style-type: none">Is the solution relevant to the business issue?Is the solution relevant to Hong Kong/China? | <ul style="list-style-type: none">How can such solution be applied to real business situations? Participants should provide 2 examples of real world commercial application for the sample technique. | 30 |
| Impact <ul style="list-style-type: none">Is the solution adding value to the business? | <ul style="list-style-type: none">The value of the two real world application examples will be measured.Suggestions on how to further fine tune the model. | 10 |
| Innovation <ul style="list-style-type: none">Does the solution introduce new ideas or methods?Does it make use of open data? | <ul style="list-style-type: none">Any application of open data or innovative algorithm to the solution to enhance the accuracy of prediction?Function include embedment into the camera | 20 |
| Technical achievement <ul style="list-style-type: none">Readiness to use (Technical accuracy)Ease of use | <ul style="list-style-type: none">Reason of the methodology used (Pros and cons for such selection)Quality of the programme coding (e.g. Object-orientation, soft-coding (feature uses “engines” that derive results based on parametric values (i.e. business rules), dynamic coding / flexibility) | 30 |
| Communication and presentation | - | 10 |

Challenge 2: Payment prediction



Background and the need of the challenge

Many companies struggle with short payable periods and could improve their working capital by detecting early on potential issues in future vendor payments.

By analysing historical payment timing data, PwC would like to predict future payment dates for outstanding vendor invoices.



Data provided

We will provide 2 files in a website once your team have signed the non-disclosure agreement:

1. Filename InvoicePayment-training.csv – which is 3 years of historical payment transactions
2. Filename InvoicePayment-evaluation.csv – which is 3 months of payment transactions that you will need to predict the actual payment date

A data dictionary is provided to you in the subsequent slide.

Challenge 2: Payment prediction

We will provide the following data dictionary for the input file:

| Column Name | Data Type | Description |
|-----------------------------|-----------|---|
| PwC_RowID | int | Unique System ID |
| BusinessTransaction | varchar | Business Transaction Type |
| CompanyCode | varchar | Company Code |
| CompanyName | varchar | Company Name |
| DocumentNo | varchar | Invoice No |
| DocumentType | varchar | Invoice Type Code |
| DocumentTypeDesc | varchar | Invoice Type Description |
| EntryDate | date | Invoice Entry Date |
| EntryTime | time | Invoice Entry Time |
| InvoiceAmount | numeric | Invoice Amount |
| InvoiceDate | date | Invoice Date |
| InvoiceDesc | varchar | Invoice Description |
| InvoiceItemDesc | varchar | Invoice Item Description (detail level of Invoice) |
| LocalCurrency | varchar | Invoice Currency |
| PaymentDate | date | Payment Date |
| PaymentDocumentNo | varchar | Payment No |
| Period | varchar | Month |
| PO_FFlag | varchar | Identifies whether a Purchase Order is matched with the Invoice |
| PO_PurchasingDocumentNumber | varchar | Invoice Corresponding Purchase Order No |
| PostingDate | date | Invoice Post Date |
| PurchasingDocumentDate | date | PO Date |
| ReferenceDocumentNo | varchar | Reference Document No |
| ReportingAmount | numeric | Reporting Amount |
| TransactionCode | varchar | Transaction Code (detail level of business transaction) |
| TransactionCodeDesc | varchar | Transaction Code Description |
| UserName | varchar | Processing User Name |
| VendorName | varchar | Vendor Name |
| VendorCountry | varchar | Vendor Country Code |
| Year | int | Year |
| PaymentDueDate | date | Payment Due Date |

Challenge 2: Payment prediction



Submission

You will need to submit the following to **data-lympic@hk.pwc.com** by 12pm of 29 January 2019:

1. Source code of your solution
2. List of library used by your solution
3. The prediction result of your solution

You have to predict the actual payment date for the invoices in the 3 last months. You will also have to predict whether these payments are going to be early, on time, or late by comparing the predicted payment date with the payment due date.

You are required to create a coma-separated CSV file with three columns including “PwC_RowID”, “PredictedPaymentDate”, and “PaymentTiming” (Early, On time or Late).

Below is a demonstration of what your file should look like:

```
PwC_RowID,PredictedPaymentDate,PaymentTiming
1,2016-10-11,Early
2,2017-01-14,Early
3,2016-07-12,Late
4,2016-09-14,On time
5,2017-10-21,On time
6,2017-11-19,Late
```

Challenge 2: Payment prediction

Round 1 evaluation: Comparison of predicted payment date with the actual payment date

We will calculate the date difference of each payment transaction, and the absolute value of the difference will then be summarised to compare with all other participating teams.

The lower the final number, the higher the ranking will be.

This is an illustrative example:

| PwC_Row | PredictedPaymentDate | PaymentTiming | ActualPaymentDate | Difference |
|---------|----------------------|---------------|-------------------|------------|
| 1 | 2016-10-11 | Early | 2016-10-12 | 1 |
| 2 | 2017-01-14 | Early | 2017-01-10 | 4 |
| 3 | 2016-07-12 | Late | 2016-07-14 | 2 |
| 4 | 2016-09-14 | On time | 2016-09-14 | 0 |
| 5 | 2017-10-21 | On time | 2017-10-28 | 7 |
| 6 | 2016-11-19 | Late | 2016-11-10 | 9 |
| | | | | |
| | | | Final score | 23 |

Challenge 2: Computer vision challenge

Round 2 evaluation: Presentation

| Evaluation criteria | Details | Points |
|---|---|--------|
| Relevance <ul style="list-style-type: none">Is the solution relevant to the business issue?Is the solution relevant to Hong Kong/China? | <ul style="list-style-type: none">How can such solution be applied to real business situations? Participants should provide 2 examples of real world commercial application for the sample technique. | 30 |
| Impact <ul style="list-style-type: none">Is the solution adding value to the business? | <ul style="list-style-type: none">The value of the two real world application examples will be measured.Suggesting what other features of the vendors / payments be collected to fine tune the model | 10 |
| Innovation <ul style="list-style-type: none">Does the solution introduce new ideas or methods?Does it make use of open data? | <ul style="list-style-type: none">Any application of open data or innovative algorithm to the solution to enhance the accuracy of prediction? | 20 |
| Technical achievement <ul style="list-style-type: none">Readiness to use (Technical accuracy)Ease of use | <ul style="list-style-type: none">Reason of the methodology used (Pros and cons for such selection)Quality of the programme coding (e.g. Object-orientation, soft-coding (feature uses “engines” that derive results based on parametric values (i.e. business rules), dynamic coding / flexibility) | 30 |
| Communication and presentation | - | 10 |



Open for any questions!

Disclaimer

By participating in the Data-lympics, you agree that, to the maximum extent permitted under applicable laws, PricewaterhouseCoopers Limited (“PwC”), other PwC firms and entities, their partners, employees, officers, agents, representatives and sub-contractors (collectively, “PwC Parties”) shall not be liable or responsible for, and shall be released from and held harmless against, all liabilities, claims, demands or actions of any kind, including but not limited for any death, accidents, injuries, losses and/or damages of any kind whatsoever to any persons and property which may arise or be sustained from or in connection with the Data-lympics.

You acknowledge and agree that PwC Parties are not responsible for any technical, hardware, software, systems, networks, websites, platforms, servers and/or equipment malfunctions or failures, or any inaccurate, incomplete or incorrect information, or any other problems, errors, disruptions or interruptions of any kind whatsoever which may arise or occur in connection with the Data-lympics.

You understand and agree that PwC Parties shall have absolute and sole discretion and right to conduct, organise, suspend, modify or terminate the Data-lympics in any manner and at any time as they see fit.

You should, and should also ensure and procure that all your team members, comply with the Data-lympics Rules and maintain the highest standard of professionalism, ethics and integrity during the Data-lympics. As such, you and your team should not attempt or take any actions to sabotage, tamper or interfere with other competing teams, the Data-lympics’s platforms or any PwC Parties’ hardware, software, systems, networks, websites, platforms, servers and/or equipment. You understand and agree that PwC Parties have absolute discretion to disqualify your team for any violation of the Data-lympics Rules or ethics/integrity standards.

You agree that any disputes on or relating to the Data-lympics shall be governed by the laws of Hong Kong, and the Hong Kong courts shall have exclusive jurisdiction over any disputes, whether contractual or non-contractual.

Thank you

pwchk.com

© 2019 PricewaterhouseCoopers Limited. All rights reserved. Not for further distribution without the permission of PwC. "PwC" refers to the Hong Kong member firm, and may sometimes refer to the PwC network. Each member firm is a separate legal entity. Please see www.pwc.com/structure for further details.