

Game Data Analysis

Tencent Data Science Intern
Take Home Challenge

Jiawei Xia | Mar 25

Content

01

Anomaly Detection

02

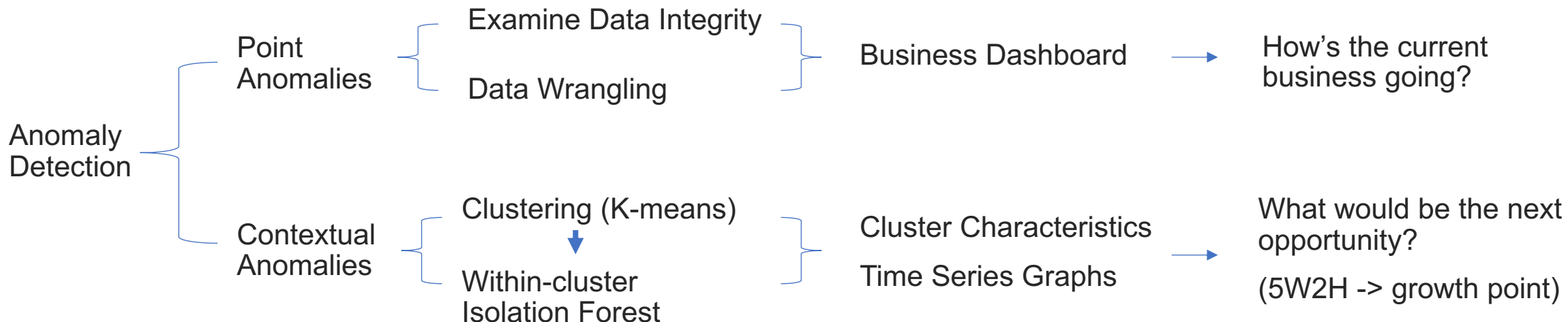
Data Science
(Clustering/Prediction)

03

Visualization/
Time Series

04

Questions Answered



Point Anomalies

Examine Data Integrity 

Data Wrangling

Data Frame

- › Head sample data
- › Shape (748264 x 14)
- › Data types
- › Missing values (0)

Actions:

- › Transform str object into datetime for “Date”

Columns

- › Date range (from 2018-03-15 to 2020-03-09)
- › Countries (239)
- › Platforms (3)
- › Key stats (min, max, etc)
- › Duplicated columns (1)

Actions:

- › Remove anomalies disobeying business rules
- › Drop the second “Date”

Rows

- › Duplicated rows (~50%)
- › Conflicting rows (in country ‘Dominica’)
- › Inconsistent country format

Actions:

- › Drop repeated rows
- › Drop conflicting rows
- › Translate zh-cn to en

Pairwise

- › Diagonal: concentrated data
- › Line shape: possible linear relationship
- › L shape & U shape: 2 possible clusters
- › E shape: 3 possible clusters

Actions:

- › Assumption for clustering k=2 or 3
- › Linear correlation

Point Anomalies

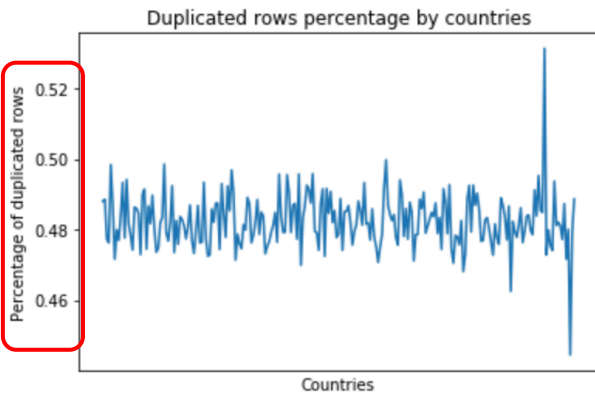
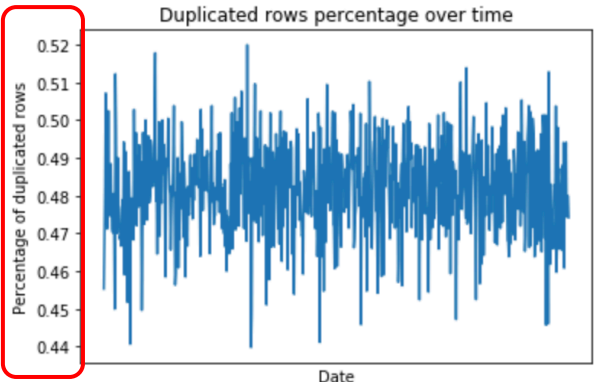
Examine Data Integrity 

Data Wrangling

Date
Date.1
Platform
Country
TRU
DAU
Items
Trans
Items Per Trans
Items per DAU
Conversion
Cash Flow
Return Customer
Time Spend Per Day(seconds)
dtype: object

Iran(伊朗) 3441
Australia(澳大利亚) 3434
圣卢西亚 3366
Israel(以色列) 3365
Iraq(伊拉克) 3363
...
厄立特里亚 529
托克劳 496
圣巴泰勒米 239
unknown(未知) 31
蒙塞拉特岛 18

object
object
object
float64
float64
float64
float64
float64
float64
float64
float64
float64
float64
float64



	TRU	DAU	Items	Trans	Items Per Trans	Items per DAU	Conversion	Cash Flow	Return Customer	Time Spend Per Day(seconds)
count	7.482640e+05	7.482640e+05	7.482640e+05	748264.000000	748264.000000	748264.000000	748264.000000	748264.000000	748264.000000	748264.000000
mean	7.399568e+05	1.816949e+04	5.089120e+04	3958.49364	31.569433	1.414330	2.902036	5093.453512	39.795054	122.605716
std	6.776595e+05	1.801799e+04	4.610241e+04	3655.62504	23.629177	1.208879	2.591666	4712.790645	24.796009	43.331318
min	6.286478e+05	1.529217e+04	4.342911e+04	3428.09000	17.580000	1.300000	2.300000	4527.010000	12.210000	-9.990000
25%	6.288102e+05	1.529677e+04	4.342911e+04	3428.09000	17.580000	1.300000	2.300000	4527.010000	28.660000	98.950000
50%	6.314452e+05	1.537393e+04	4.352124e+04	3433.85000	28.090000	1.320000	2.560000	4532.770000	39.430000	118.180000
75%	6.574735e+05	1.617663e+04	4.466021e+04	3522.53000	36.080000	1.390000	3.000000	4611.080000	48.920000	139.370000
max	2.677675e+07	2.141935e+06	4.387412e+06	434846.58000	1744.870000	444.240000	232.630000	819825.280000	3912.090000	1390.060000



Point Anomalies

Examine Data Integrity

Data Wrangling

Date

Date.1

Platform

Country

TRU

DAU

Items

Trans

Items Per Trans

Items per DAU

Conversion

Cash Flow

Return Customer

Time Spend Per Day(seconds)

dtype: object

Drop

Transform

object
object
object
float64
float64
float64
float64
float64
float64
float64
float64
float64
float64
float64

Iran(伊朗)

Australia(澳大利亚)

圣卢西亚

Israel(以色列)

Iraq(伊拉克)

厄立特里亚

托克劳

圣巴泰勒米

unknown(未知)

蒙塞拉特岛

3441

3434

3366

...

529

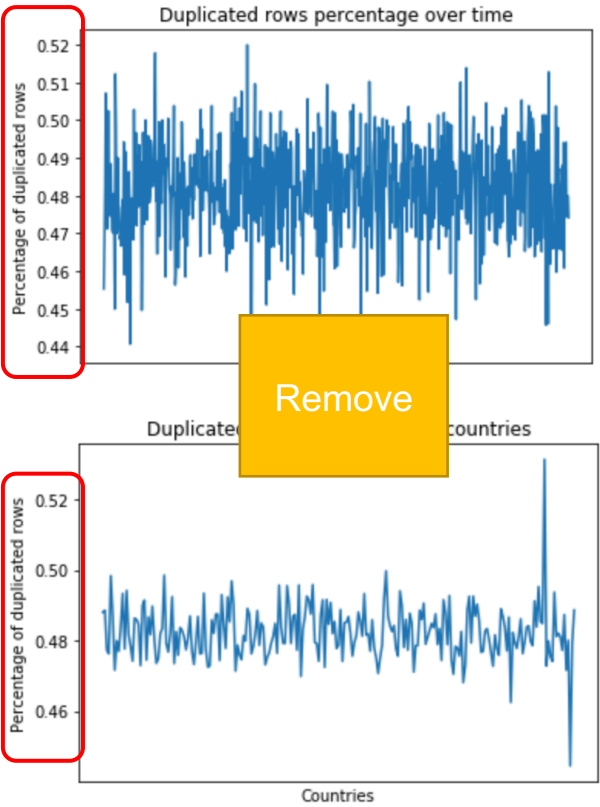
496

239

31

18

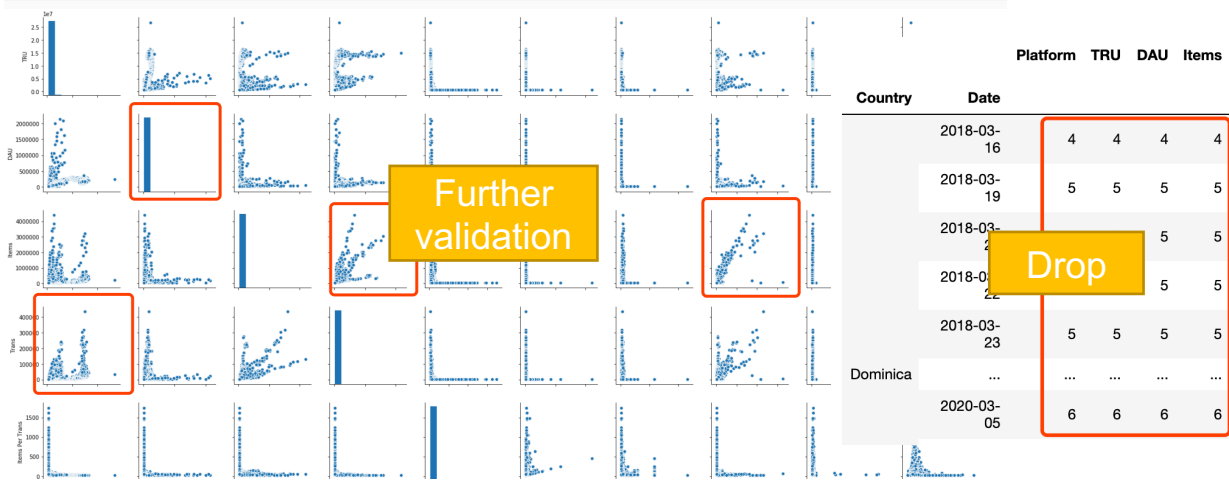
Translate



	TRU	DAU	Items	Trans	Items Per Trans	Items per DAU	Conversion	Cash Flow	Return Customer	Time Spend Per Day(seconds)
count	7.482640e+05	7.482640e+05	7.482640e+05	748264.000000	748264.000000	748264.000000	748264.000000	748264.000000	748264.000000	748264.000000
mean	7.399568e+05	1.816949e+04	5.089120e+04	3958.49364	31.569433	1.414330	2.902036	5093.453512	39.795054	122.605716
std	6.776595e+05	1.801799e+04	4.610241e+04	3655.62504	23.629177	1.208879	2.591666	4712.790645	24.796009	43.331318
min	6.286478e+05	1.529217e+04	4.342911e+04	3428.09000	17.580000	1.300000	2.300000	4532.770000	39.430000	-9.990000
25%	6.288102e+05	1.529677e+04	4.342911e+04	3428.09000	17.580000	1.300000	2.300000	4532.770000	39.430000	98.950000
50%	6.314452e+05	1.537393e+04	4.352124e+04	3433.85000	28.090000	1.320000	2.560000	4611.080000	48.920000	118.180000
75%	6.574735e+05	1.537393e+04	4.352124e+04	3433.85000	28.090000	1.320000	2.560000	4611.080000	48.920000	139.370000
max	2.677675e+07	2.677675e+07	2.677675e+07	4846.58000	1744.870000	444.240000	232.630000	819825.280000	3912.090000	1390.060000

Further validation

Remove



Business Dashboard

Tencent

Business Dashboard

Tencent

Contextual Anomalies

K-Means Clustering 

Within-cluster Isolation Forest

Data Wrangling



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Find the Best K



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Fit and Evaluate Model



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Examine Clusters

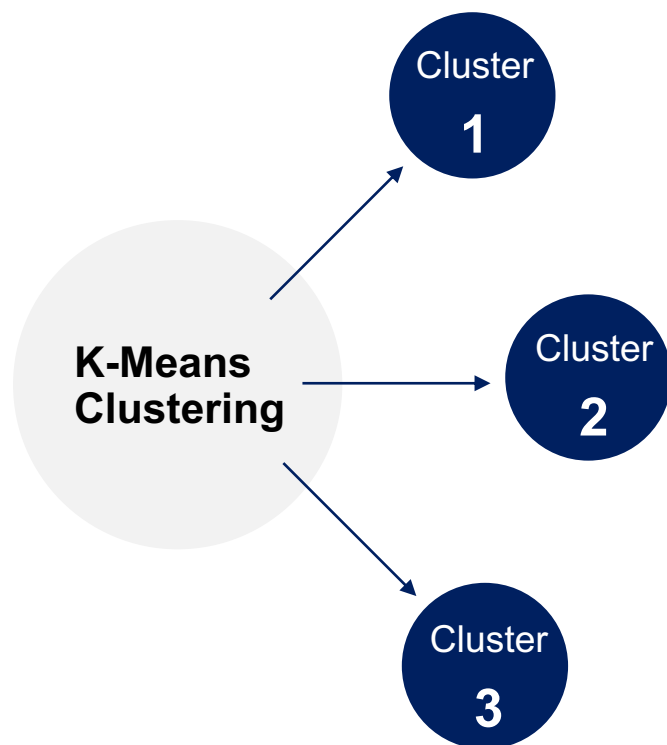


- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Contextual Anomalies

K-Means Clustering ✓

Within-cluster Isolation Forest ✓



Cluster Characteristics

Cluster Time Series Analysis

Anamoly Time Series Analysis

Tencent

Conclusion



1st Issues

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.



2nd Issues

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.



3rd Issues

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

The background of the slide is a blue-tinted photograph of two men in an office. The man on the left is wearing glasses and a dark t-shirt, while the man on the right is wearing a light-colored hoodie. They are both smiling and looking at a laptop screen. A white rectangular border is centered over the image, containing the text 'THANK YOU'.

THANK YOU

Content

01

Lorem ipsum dolor
sit amet,
consectetuer
adipiscing elit.

02

Lorem ipsum dolor
sit amet,
consectetuer
adipiscing elit.

03

Lorem ipsum dolor
sit amet,
consectetuer
adipiscing elit.

04

Lorem ipsum dolor
sit amet,
consectetuer
adipiscing elit.

Project status report



Status 1



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.



Status 2



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.



Status 3



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

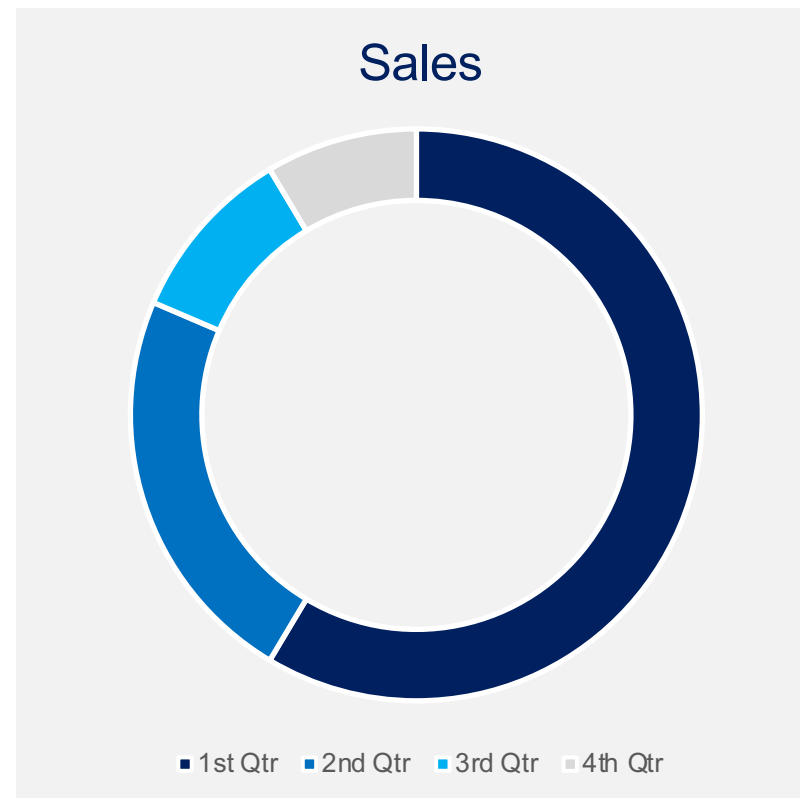
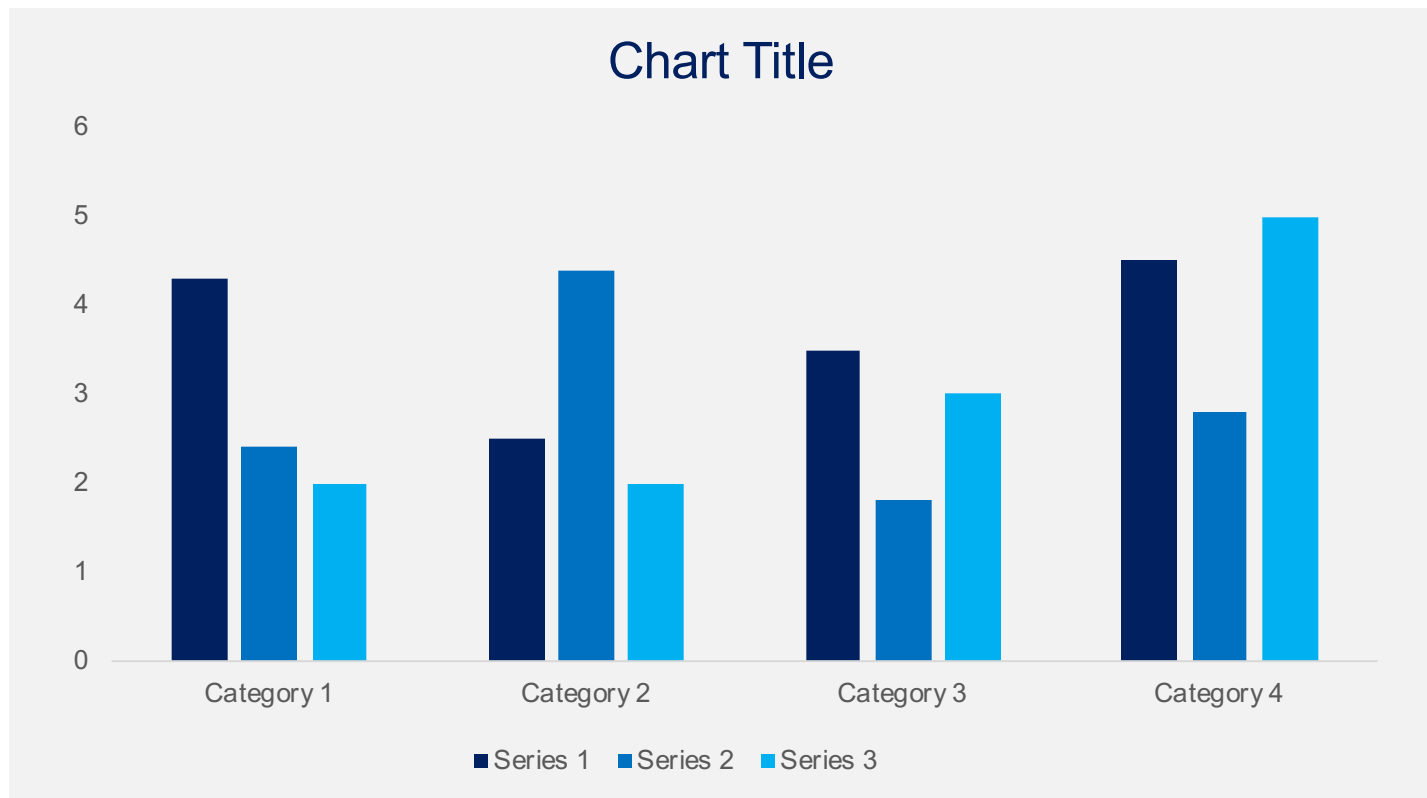


Status 4



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Data report



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas porttitor congue massa.

Risks 1



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Risks 2



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Risks 3



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.
- › Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci.

Summary

Project	Progress	Status	Team	Comments
Project 1	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum dolor sit amet.
Project 2	Lorem ipsum	Lorem ipsum	Lorem ipsum	Consectetuer adipiscing elit.
Project 3	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum dolor sit amet.
Project 4	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum dolor sit amet.
Project 5	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum dolor sit amet.
Project 6	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum dolor sit amet.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Solutions offer



Next steps

August 2019 plans

WEEK 1



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

WEEK 2



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

WEEK 3



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

WEEK 4



- › Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- › Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.