|  |  |
| --- | --- |
|  | |
| Data Analysis and Visualisation Project  Data Vis and Comms CA – Report | |
| Module code : B8IT107 | |
| Ciaran Finnegan  Student No : 10524150  26/03/2020 |  |
|  |  |

Table of Contents

1 Project Overview 3

1.1 High Level Description – Software Quality Assurance Dashboards 3

1.2 Environment Assumptions 4

1.3 Project Execution Instructions 4

1.4 Data Privacy Considerations 5

2 Part 1: Understanding The Business Drivers 6

2.1 Business Objective 6

2.2 Subject Area for Analysis 6

2.3 Primary Goals and Objectives for the Data Visualisations 7

2.4 Building Blocks for the Quality Assurance Dashboards 8

2.5 Key Stakeholders 11

3 Part 2: Data Preparation 12

3.1 Collating Excel Data 12

3.2 Preparing the Production Defect Data Extract 12

4 Part 3: Data Ingestion and Manipulation 13

4.1 Data Ingestion Using PowerBI – Query Builder 13

4.2 Data Manipulation in Power BI – Post Query Editor 13

5 Part 4 : Data Analysis / Visualisation 14

5.1 Dashboard One Design – Defect Analysis Dashboard 14

5.2 Dashboard Two Design – Defect Priority Dashboard 22

5.3 Dashboard Three Design – Test Automation Dashboard 26

6 Conclusions 30

6.1 Interpretations from the Quality Assurance Visualisations 30

6.2 Challenges in the Quality Assurance Data Visualisations 30

7 Appendices 31

7.1 Appendix 1 – Appendix Title 31

8 References 32

8.1 Reference One 32

# Project Overview

## High Level Description – Software Quality Assurance Dashboards

This document covers the design, implementation and observations on all parts of the March 2020 CA for the Data Visualisation and Communications module (BTIT107).

This project is based on real world requirements and data to provide a graphical view on the effectiveness of the Quality Assurance (QA) process within the development and Production deployment lifecycle for my company.

The business domain for this project can be characterised as follows;

* This is a software engineering company that develops an enterprise level product for the Financial Services industry.
* Products are typically deployed on client site within their local Production environment (not cloud based).
* The client base is approximately 100+ customers. There is a wide geographical spread of clients.
* The ‘product’ is built up from multiple components, which are developed and tested across multiple regions.
* A new version of the product is released every six months, and deployments are facilitated through an external consultancy function. This business unit records all defects raised during the lifetime of the product into a third party tool called JIRA.

I have taken actual data on defects raised during product development and built dashboards that combine this data with Production raised defects. The intention is to create a more refined reporting process on QA effectiveness, both in general and broken down over time/location/product component.

<screenshots of dashboards>

## Environment Assumptions

Reporting dashboards were developed and presented in PowerBI.

A number of EXCEL worksheets contained the source information on defects raised during the Product development lifecycle. These had been built up over a period of 3+ years and were formatted into identical EXCEL worksheets.

The Product defect data was a .*csv* file of 2000+ records, extracted from the production defect tracking system, which is a third party tool called Atlassian JIRA.

The underlying data preparation and manipulation processes for the dashboard visualisations are described in

The output of this reporting process is described in Section 5 of this document.

To summarise, the key software components used in development of this project were;

* PowerBI
* Microsoft Excel
* Notepad++ v7.5.8 (for data views)

## Project Execution Instructions

Building the PowerBI dashboards was a process of repeated data manipulation and processing to provide the final dataset suitable for the visualisation charts I developed.

This resultant .pbix file has been included with the CA submission and can be loaded into PowerBI;

* ***<file name….>.pbix***

## Data Privacy Considerations

This project is built with real world data collated within my company over approximately the last three years.

I felt it was a suitable candidate for the Data Visualisation and Communication CA, but I also intend to put it to use in real life and improve the standard of metrics that are employed to show QA effectiveness.

Obviously when dealing with real data there is a need to be aware of company and regulatory data policies.

To address these issues, I took the following steps with the product defect data;

1. Client names and mnemonics are captured by the JIRA defect tracking system for defects in production. These have all been removed from the ***.csv*** JIRA extract.
2. Similarly, names of anyone adding or tracking a defect ‘ticket’ have been removed, regardless of whether the individuals are internal employees or customers.
3. The product version numbers have been re-sequenced to obscure the actual release descriptions.
4. Product components and office locations have been altered, although the number of components and development locations reflects the real world set up.

For internal office use in the near future I will revert some of the above data changes, made for the CA, so that I can employee the dashboard in a current work setting.

My company has an enterprise wide PowerBI licence and I will be using the online version for report generation and presentation.

# Part 1: Understanding The Business Drivers

## Business Objective

Testing is an essential part of the Software Development Lifecycle. However, how do you know if the investment in QA is being effective? How can you measure the impact that process changes are having over time?

In my company the move in the last five years to an AGILE framework for software product development has coincided with additional investment in formal test engineers and new test automation technology.

The entire concept of ‘Quality Assurance’ has taken on a ‘whole team’ methodology, not just measured by how effective the testers are at finding bugs but by how effective the coders are at removing errors before the tester or the client find them.

Are these changes having a positive impact on product quality, and by extension customer satisfaction? Can we demonstrate the effectiveness (or lack of success) of these changes in a manner that is easily understood by all stakeholder.

This is the objective of the dashboard visualisations in this project.

## Subject Area for Analysis

A brief review of ISTQB recommended practices (the commonly adopted standard for QA process) suggests a range of metrics to use to measure QA effectiveness.

In this assignment I have chosen to focus on an analysis of*the ratio of defects found during the development phase versus the amount of defects subsequently uncovered by the client*.

## Primary Goals and Objectives for the Data Visualisations

The cost of addressing production defects on client site is considerably higher than during development so the majority of bugs should be captured prior to product release.

However, the priority of a defect is also a key characteristic of the metrics. In commercial software spending time to fix every cosmetic defect may delay a release to market, which could result in loss of market share. For non-critical software, a sensible balance should be established to determine that the bulk of post release defects, although undesirable, are not responsible for serious processing flaws in a client production environment.

My dashboards are a means to graphically analyse that the company is striking the right proportions in terms of defect capture, across components and development locations.

## Building Blocks for the Quality Assurance Dashboards

In recent history, each product release made by my company has produced a set of QA metrics for the formal Release Note documentation.

This has included Excel spreadsheets with the following data;

* Release Number
* Release Date
* Defects Found per Component (a typical release comprises nn separate components, some of which are developed in different locations and integrated before release).

Other Excel spreadsheet is updated on a less frequent basis with the following data with to frame historical release information with client production data;

* Post Release Period 1
* Post Release Period 2

These date period reflect two consecutive deployment windows for each release. A key business objective is to determine the frequency/volume of defects raised by clients. The purpose of these Release Periods is to;

1. Determine if an unexpectedly high volume of defects are found early in the first post release phase, possibly indicating that the QA process was ineffective at capturing defects easily spotted by clients. Experience has shown that the first four-six months after a product release is the key timeframe.
2. User the second post release phase to effectively establish a ‘cut-off’ point for client defects. Practical experience has shown that with a large client base multiple clients will start to report the same defects, therefore skewing the metrics. In practice I have found that limiting the second period to approximately 8 months covered the vast majority of clients taking new product versions, while avoiding duplication of issues being captured across the client base.

These dates are captured on a release by release basis. For the purposes of the CA I have provided the date periods but, for the sake of brevity, I have not elaborated on the historical reason for the date ranges themselves.

Previous processes involved laborious extraction of multiple spreadsheets, exported from the JIRA tool, to build up the production defect data.

Taking advantage of the functionality within PowerBI I have chosen to streamline the process by reading a full .csv extract, generated by JIRA, into PowerBI and build up a meaningful relationship between the data to produce the visualisations in this CA.

### Charting Development vs Production Metrics

Dashboard One focuses on defect volumes, and a graphical comparison of development numbers versus production

Dashboard Two provides additional information on…

### KPIs

To start with a definition;

A **Key Performance Indicator** (KPI) is a measure to provide managers with the most important performance information required to enable them, or their stakeholders, understand the performance of the business.

For this project, I have identified the following KPIs to assist stakeholders in understanding the data visualisations;

**One KPI**

KPI 1: Measures ..

**Two KPI**

KPI 2: Measures ..

### Geographical and Component Data – Pinpoint Focus Areas

This project also aims to provide a graphical analysis of the ratio of defects based on component and location. Both of these metrics help identify a particular team unit responsible for the element of product delivery.

This is not a ‘name and shame’ exercise, but rather an attempt to look at areas in the business that could possibly benefit from further investment in QA training or processes.

### Test Automation – is it working?

Test automation is an integral part of any modern software development and test process. However, it can be expensive and challenging to deploy and business stakeholder will expect to see a return on this type of investment.

Dashboard Two has included a visual to make the case for the ongoing benefits of test automation.

(In this CA the test automation is Selenium based UI and API level tests to replicate/replace the traditional manual test process).

## Key Stakeholders

The reports produced by the work in this CA will ultimately lead to a more sophisticated and distributable reporting model for QA activity in my company.

The key stakeholders for the type of reporting produced in this project are therefore;

* **Senior Management.** Those that ask the question; can you prove that your QA processes deliver value?
* **Mid-level Management.** The people like me that want to see first-hand evidence that process changes we championed within our teams are actually working. Do we have an impressive way to convey this to higher levels of management?
* **Partners.** Those that deploy our software at client site need assurances that we are working to be proactive and continually trying to innovate with product quality.
* **Employees.** It team members are part of an effective delivery group, then can we measure this and reward appropriately?

# Part 2: Data Preparation

## Collating Excel Data

The ...

## Preparing the Production Defect Data Extract

The .

# Part 3: Data Ingestion and Manipulation

## Data Ingestion Using PowerBI – Query Builder

The ..

## Data Manipulation in Power BI – Post Query Editor

The..;

# Part 4 : Data Analysis / Visualisation

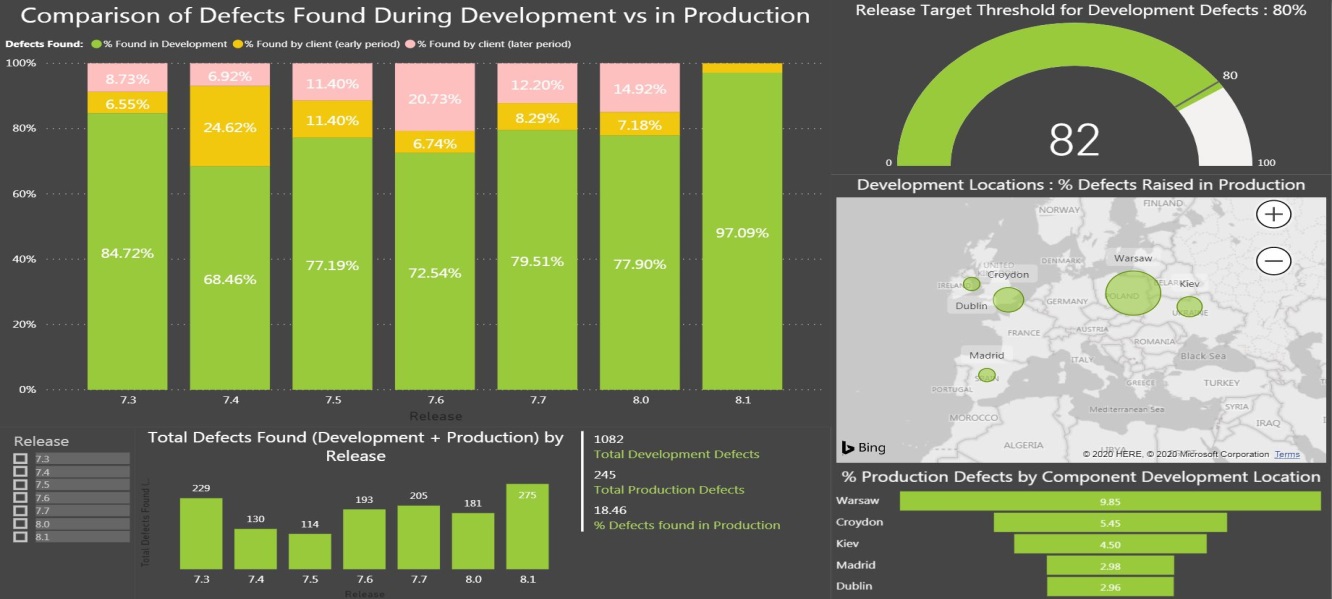
## Dashboard One Design – Defect Analysis Dashboard

The purpose of this dashboard is to show the ratio of software defects found during the development phase, versus those found in production environment by clients/external consultants.

**Dashboard 1: Defect analysis Dashboard**

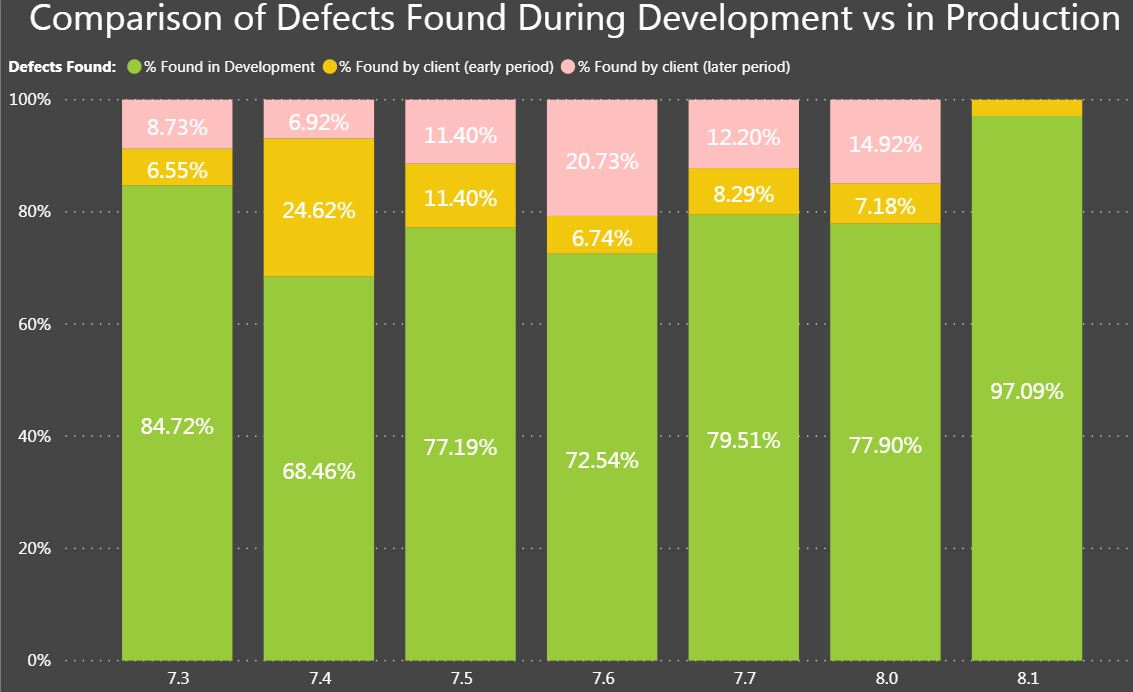
|  |  |
| --- | --- |
| **Dashboard Component** | **Visualisation (reading clockwise from top)** |
| **Tile 1** | **Comparison of Development vs Production Defects** |
| **Tile 2** | **Release Target Thresholds.** |
| **Tile 3** | **% Production Defects by Development Location (Map)** |
| **Tile 4** | **% Production Defects by Development Locations (Funnel)** |
| **Tile 5** | **High Level Metrics on Development Defects vs Production** |
| **Tile 6** | **Total Defects Found per Release** |

The first dashboard is a collection of visuals intended to show how effective the company QA process has been, over the period of the most recent product releases, in detecting defects before being shipped to clients.



### Tile 1.1: Comparison of % Defects Found in Development vs Production

This tile is a 100% stacked bar chart to show the ratio, in percentages, of the volume of defects found during the software development lifecycle of recent releases, versus those found in production on client site.



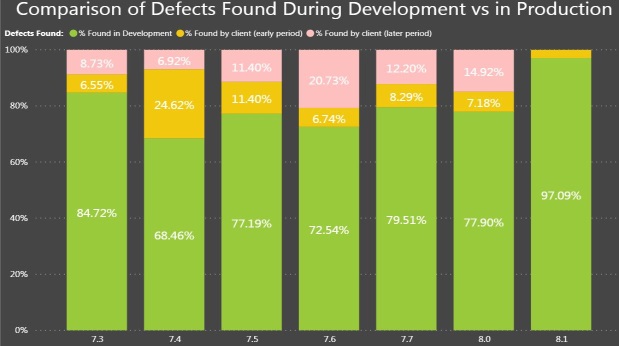
#### Purpose of Visualisation

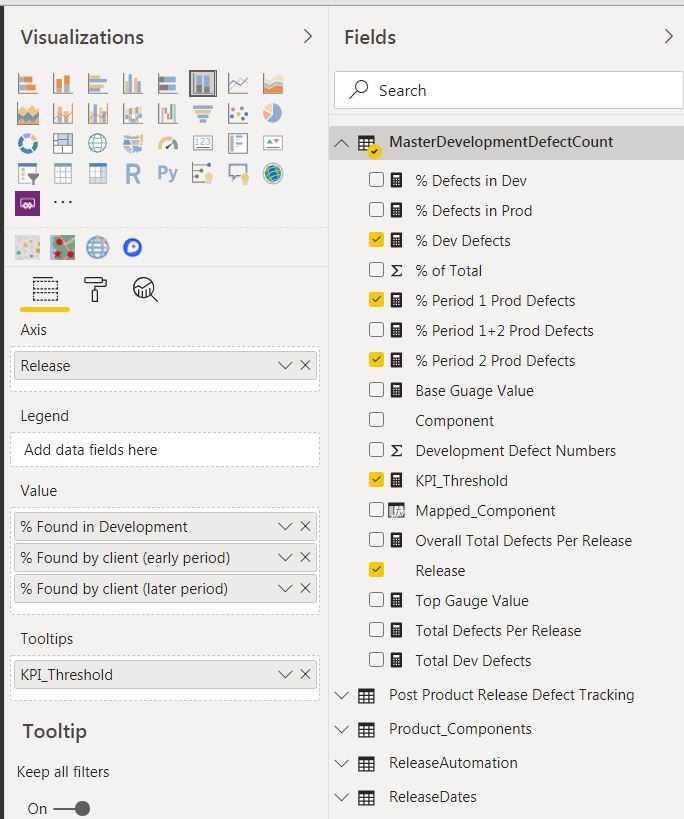
This tile provides a visual representation of how many defects are being caught by the company QA process during product development.

#### Visualisation Process

This graph is built solely from the *‘MasterDevelopmentDefectCount’* data source. The import of this data is described in more detail in Section 3 and Section 4 of this document.

Section 4 also describes the individual measures created for this dashboard in more detail.





The ‘Release’ axis is a field read directly from the Excel file into the data source but the values for the percentage of the development and production defects are calculated measures.

* % Dev Defects – this generates the initial ‘green’ bar.
* % Period 1 Prod Defects – this generates the ‘amber’ stacked bar.
* % Period 2 Prod Defects – this generates the ‘pink’ stacked bar.

A tooltip is provide for a hover over, to advise the audience of the dashboard that 80% is the ideal minimum threshold for the proportion of development defects.

#### Why This Type of Visualisation?

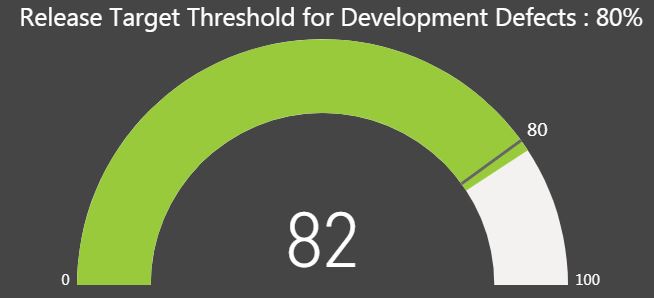
The 100% Bar Chart is used because we are looking at the entirety of defects in the development and production periods and comparing relative proportions.

#### Relation To Other Tiles in Dashboard.

Although a Slider control is provided on the dashboard, a user can also click on a given release bar in the visualisation and the other tiles will filter on that given release.

### Tile 1.2: <Tile Two Description>

This tile is a relatively simple gauge visual used to emphasise when the development defect detection rate has exceeded the ideal minimum threshold – 80% of defects captured during the development process.



#### Purpose of Visualisation

The user of the dashboard will initially see a default gauge value that combines metrics for all releases, ON how well the QA process is working at capturing bugs during the development phase.

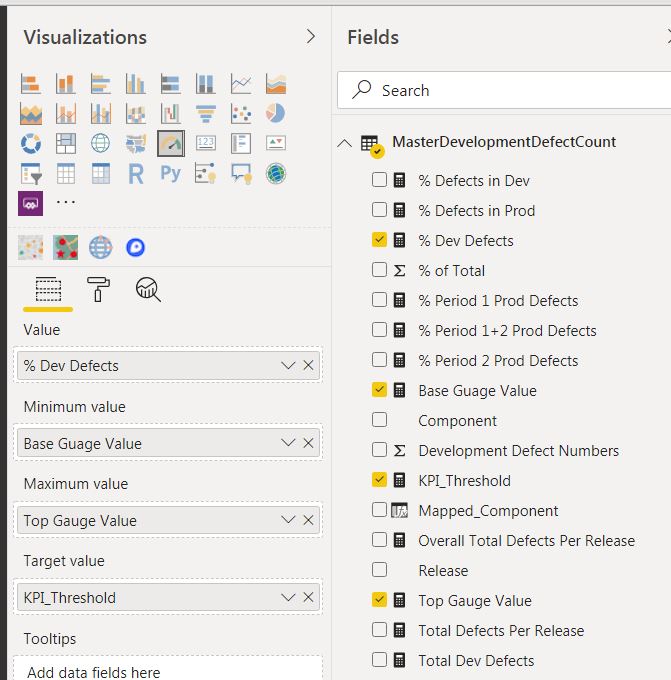
The visualisation will change focus on specific product release(s) based on user selection elsewhere on the dashboard.

#### Visualisation Process

As with Tile 1 only the *‘MasterDevelopmentDefectCount’* data source is used and the visualisation is really driven by just the measure ‘% Dev Defects’

<screenshots>





The upper and lower bound values of the guage are hard coded to reflect a 0 – 100% range, as is the KPI target of 805.

#### Why This Type of Visualisation?

This is a simple gauge to show if the development QA process is hitting the desired target for defect detection.

It can focus on one or more releases, or all releases in the project period.

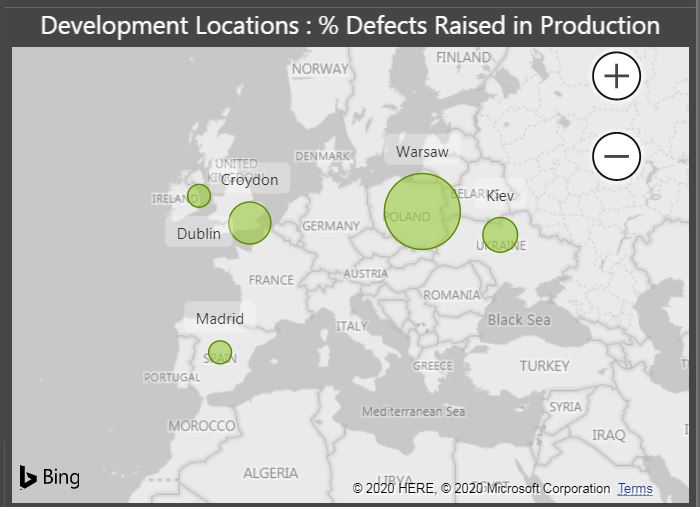
#### Relation To Other Tiles in Dashboard.

The KPI target value is statis but the valiue for the actual percentage will alter depending on the ‘Release’ selected in the other bar charts on the dashboard.

### Tile 1.3: % of Production Defects by Development Locations

This tile shows a map of company offices where software development and QA activity takes place.

The ‘bubble’ size reflect a relative proportion of defects found in production for components developed in these locations.



#### Purpose of Visualisation

This is a relatively simple, and little crude, view of the offices generating software components that have recorded defects in production on client site.

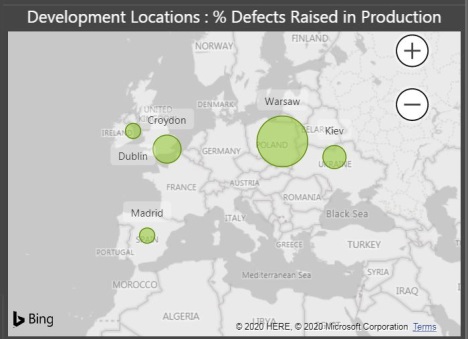
As mentioned elsewhere in this report, it is not about assigning blame but looking if there is a need to review QA processes in any given location.

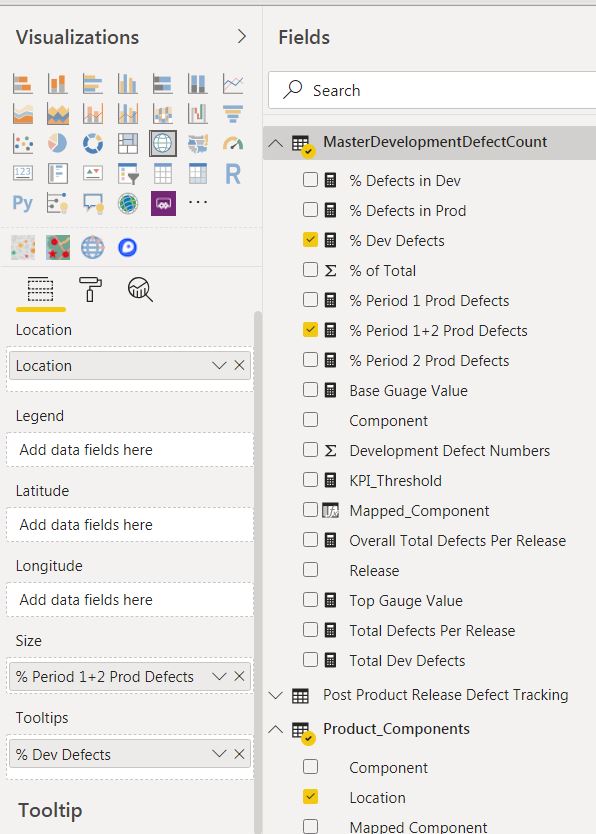
Warsaw might be a case for QA process review but in reality the spread of production defects contributed by the different offices is relatively uniform.

#### Visualisation Process

The map visualisation is driven off a relationship link from the PowerBI *MasterDevelopmentDefectCount’* data source to the reference data for ‘Product\_Components’.

<screenshot>





Defects numbers are linked to city offices through the locations for each component development.

In the company, software development for a particular component is kept within a distinct geographical location.

#### Why This Type of Visualisation?

As this is an assessment of work from different software development sites within the company it was logical to represent this data in a map visual

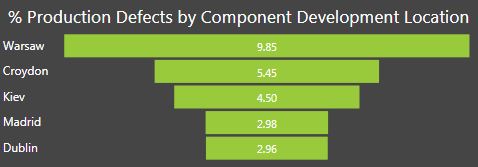
#### Relation To Other Tiles in Dashboard.

Selecting different ‘Release’ version(s), either through the Slider or by selecting a give bar in one of the charts, will update the Map graph and associated ‘buble’ for each office.

### Tile 1.4: : % of Production Defects by Development Component Locations

This tile is a compliment to the map tile (Tile 1.3) to show the relative percentage of defects found in production based on component development location.

(Each component is only developed in one location, although a location may develop multiple components).

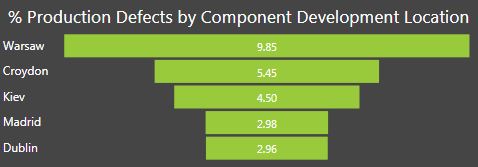


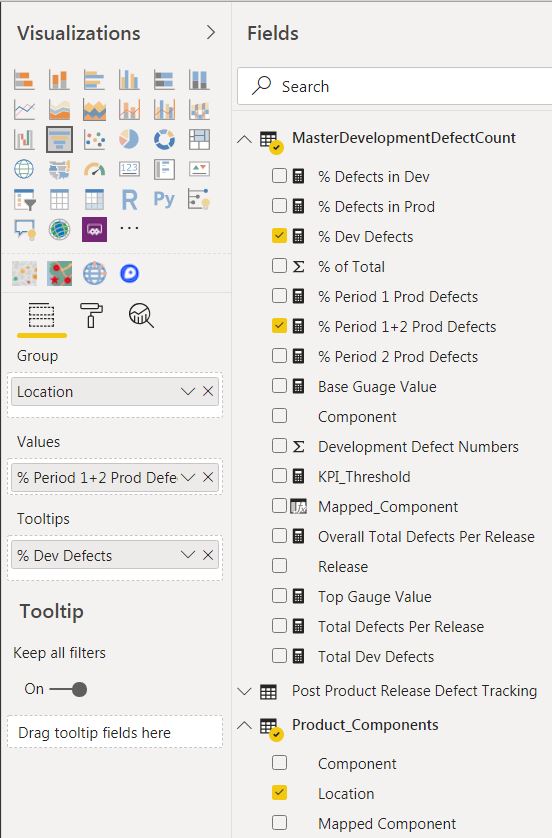
#### Purpose of Visualisation

Provide a ‘funnel’ visual to show relative percentage of defects found in production for software components developed in a given company office.

#### Visualisation Process

The tables and measures follow the same principle as those used in the map tile (Tile 1.3).





#### Why This Type of Visualisation?

It is a simple graphical aid to emphasise the date represented in the ‘bubbles’ in the tile above it in the dashboard.

#### Relation To Other Tiles in Dashboard.

Selecting different ‘Release’ version(s), either through the Slider or by selecting a give bar in one of the charts, will update the horizontal columns in the ‘funnel’.

### Tile 1.5: High Level Defect Numbers – Development and Production

This is a simple of set of numbers showing the overall number of defects found per Release(s), and how that compares with production numbers.



The numbers default, when the dashboard loads, to the over defect metrics for all software product releases in the project period.

#### Purpose of Visualisation

This tile provides a set of three numbers to provide a number to compliment the visuals in the bar charts in the dashboard.

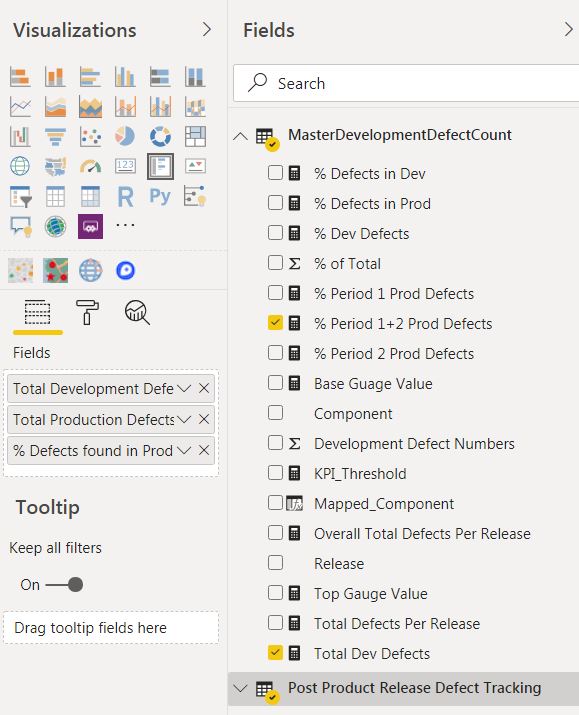
Tile 1.1 shows percentage ratios of defects, while this tile shows the actual numbers of defects driving the percentage calculations.

#### Visualisation Process

This tile directly reads the following measures from both the PowerBI *‘MasterDevelopmentDefectCount’* and ‘*Post Product Release Defect Tracking*’ data sources:

* Total Dev Defects
* Total Production Period Defects
* % Period 1 + 2 Prod Defects





The measures have been renamed on the actual tile for the purposes of presentation.

#### Why This Type of Visualisation?

This ‘visualisation’ provides some concrete numerical values to compliment the bar charts on the visualisation.

#### Relation To Other Tiles in Dashboard.

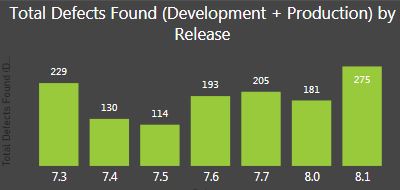
Selecting different ‘Release’ version(s), either through the Slider or by selecting a give bar in one of the charts, will update the numnerical values in this Mult-Row Card visual.

### Tile 1.6: Defect Numbers Found Per Release

This tile provides a quick visual on the scale of defects found in each product release. It combines the numbers found by the company QA process during software development, along with those found after the product shipped to client site.

(The production numbers are deliberately truncated after a certain period to avoid double counting across the company client base – see Section 2.4 in this document for details).

<screenshot>



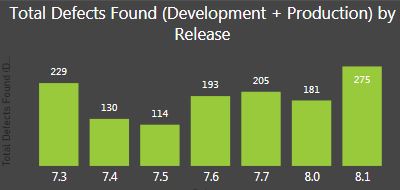
#### Purpose of Visualisation

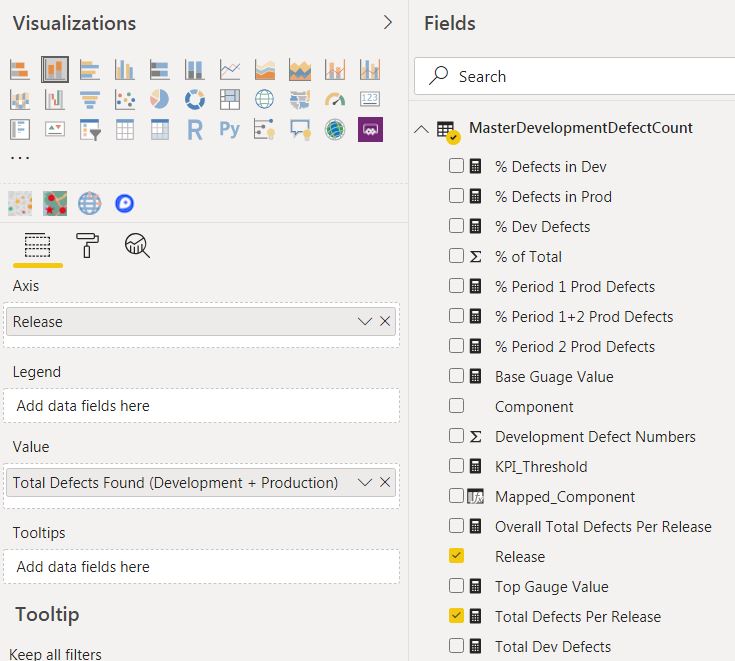
This is a straightforward bar chart visualisation of the numbers of defects found for a given release over the development and production deployment ‘lifetime’.

#### Visualisation Process

The axis on the bar chart is determined by the ‘Release’ entry in the *‘MasterDevelopmentDefectCount’* PowerBI data source.

<screenshot>





The size of the bars for each release is determined by the value in the measure : ‘Total Defects Per Release’.

The measure has been renamed on the visualisation to ‘Total Defects Found (Development + Production)’ for the purposes of presentation on the dashboard.

#### Why This Type of Visualisation?

This bar chart compliments the ratio comparision in Tile 1.1 and is an effective way to show the numbers of defects in a series of product releases.

#### Relation To Other Tiles in Dashboard.

Selecting different ‘Release’ version(s), either through the Slider or by selecting a given bar in Tile 1.1, will update this bar chart and highlight the given selected ‘Release’.

## Dashboard Two Design – Defect Priority Dashboard

The purpose of this dashboard is to provide insight into the severity of the defects raised in production environments by clients/consultants.

**Dashboard 2: Defect Priority Dashboard**

|  |  |
| --- | --- |
| **Dashboard Component** | **Visualisation (reading clockwise from top)** |
| **Tile 1** | **Production Defects by Priority** |
| **Tile 2** | **Metrics on Defect Volumes by Priority** |
| **Tile 3** | **Defect Breakdown by Component in Development** |
| **Tile 4** | **Component Mix for Defects Found in Production** |

The second dashboard is intended to show that the significant majority of defects raised by clients/consultants in production are Priority 3 or lower.

Visuals on the dashboard are also intended to show the spread of production defects across various product components and how this ratio compares to the mixture of component defects found during the development phase.

### Tile 2.1: <Tile Two Description>

This tile..

<screenshot>

#### Purpose of Visualisation

The..

#### Visualisation Process

The..

<screenshot>

#### Why This Type of Visualisation?

The..

#### Relation To Other Tiles in Dashboard.

The..

<screenshot?>

### Tile 2.2: <Tile Two Description>

This tile..

<screenshot>

#### Purpose of Visualisation

The..

#### Visualisation Process

The..

<screenshot>

#### Why This Type of Visualisation?

The..

#### Relation To Other Tiles in Dashboard.

The..

<screenshot?>

### Tile 2.3: <Tile Three Description>

This tile..

<screenshot>

#### Purpose of Visualisation

The..

#### Visualisation Process

The..

<screenshot>

#### Why This Type of Visualisation?

The..

#### Relation To Other Tiles in Dashboard.

The..

<screenshot?>

## Dashboard Three Design – Test Automation Dashboard

The purpose of this dashboard is to show the impact that increased investment in QA test automation technologies is having on defect trends, both during development and on the numbers of defects being found in production.

**Dashboard 3: Test Automation Dashboard**

|  |  |
| --- | --- |
| **Dashboard Component** | **Visualisation (reading clockwise from top)** |
| **Tile 1** | **% Increase in Test Automation vs % Production Defects** |
| **Tile 2** | **High Level Test Case Automation Metrics** |
| **Tile 3** | **High Level Development Defect Metrics (Table)** |
| **Tile 4** | **High Level Development Defect Metrics (Area)** |
| **Tile 5** | **Increase in Test Automation Numbers vs Production Defect Numbers.** |

The third dashboard shows an overlap of the increase in test automation over time, for each recent product release, and the change in defects reported. There are slight, but not spectacular improvements to be seen in the visuals.

The metrics for 8.1 are based on relatively little data, as this release has not had much time to penetrate the marketplace.

### Tile 2.1: <Tile Two Description>

This tile..

<screenshot>

#### Purpose of Visualisation

The..

#### Visualisation Process

The..

<screenshot>

#### Why This Type of Visualisation?

The..

#### Relation To Other Tiles in Dashboard.

The..

<screenshot?>

### Tile 2.2: <Tile Two Description>

This tile..

<screenshot>

#### Purpose of Visualisation

The..

#### Visualisation Process

The..

<screenshot>

#### Why This Type of Visualisation?

The..

#### Relation To Other Tiles in Dashboard.

The..

<screenshot?>

# Conclusions

## Interpretations from the Quality Assurance Visualisations

In the..

## Challenges in the Quality Assurance Data Visualisations

In the..

# Appendices

## Appendix 1 – Appendix Title

In the

# References

## Reference One

The article..