High Level Design (HLD)

Data Visualization of Bird Strikes between 2000-2011

Sudarshan Chauhan

Contents

1. Introduction

1.1 What is Low Level Design Document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the House Price Prediction dashboard. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

1.3 Project introduction

A bird strike is strictly defined as a collision between a bird and an aircraft which is in flight or on a take-off or landing roll. The term is often expanded to cover other wildlife strikes - with bats or ground animals. Bird Strike is common and can be a significant threat to aircraft safety. For smaller aircraft, significant damage may be caused to the aircraft structure and all aircraft, especially jet-engine ones, are vulnerable to the loss of thrust which can follow the ingestion of birds into engine air intakes. This has resulted in several fatal accidents. Bird strikes may occur during any phase of flight, but are most likely during the take-off, initial climb, approach and landing phases due to the greater numbers of birds in flight at lower levels. To have a closer look the following document visually depicts the data collected on Bird Strikes by FAA between 2000-2011.

2. Problem Statement

The goal of this project is to analyse the bird strike incidents happened between 2000-2011. To achieve the goal, we used a data set that is collected by FAA during 2000-2011. The objective of the project is to perform data visualization techniques to understand insights of the data. This project aims apply various Business Intelligence tools such as Tableau or Power BI to get a visual understanding of the data.

3. Dataset Information

Record Id: Unique Id for the incidents

Aircraft Type: Type of aircraft

Aircraft Name: Name of aircraft

Altitude Bin: Altitude of aircraft <1000 and >1000

Aircraft make Model: Model number of aircraft

Wildlife number struck: Number of wildlife struck in the aircraft

Wildlife number struck actual: Actual no. of wildlife struck

Effect impact to flight: Effect of impact on flight

Flight Date: Date of flight

Effect indicated damage: Whether damage is caused to aircraft or not

Aircraft no. of Engines: No. of engines in aircraft

Aircraft Airline Operator: Name of airline operator

Origin State: Origin state for the aircraft

When phase of flight: Phase of flight

Conditions Precipitation:

Remains of Wild Life Collected: Whether wild life remains collected or not

Remains of Wild life send to Smithsonian: Whether wild life remains sent toSmithsonian

Wildlife Size: Size of the wildlife

Conditions Sky: Condition of the sky

Wildlife Species: Species of the wild life

Pilot warned of birds or wildlife: Whether the pilot is warned about the wildlife

Cost Total: Total cost incurred due to bird strike incidents

Feet above ground: Height of the plane during the incidents

Number of people injured: No. of people injured during the incidents

Is Aircraft large: whether the aircraft is large or not

4. Architecture

4.1 Power BI Architecture

Power BI is a business suite that includes several technologies that work together. To deliver outstanding business intelligence solutions, Microsoft Power BI technology consists of a group of components such as:

- Power Query (for data mash-up and transformation)
- Power BI Desktop (a companion development tool)
- Power BI Mobile (for Android, iOS, Windows phones)
- Power Pivot (for in-memory tabular data modeling)
- Power View (for viewing data visualizations)
- Power Map (for visualizing 3D geo-spatial data)
- Power Q&A (for natural language Q&A)

In simple terms, a Power BI user takes data from various data sources such as files, Azure source, online services, Direct Query or gateway sources. Then, they work with that data on a client development tool such as Power BI Desktop. Here, the imported data is cleaned and transformed according to the user's needs.

Once the data is transformed and formatted, it is ready to use in making visualizations in a report. A report is a collection of visualizations like graphs, charts, tables, filters, and slicers.

Moving on to the chain of processes, you can publish the reports created in Power BI desktop on two kinds of platforms; Power BI Service and Power BI Report Server.

Power BI Service is a cloud-based public platform whereas Power BI Report Server is an on-premise platform protected by firewall security.
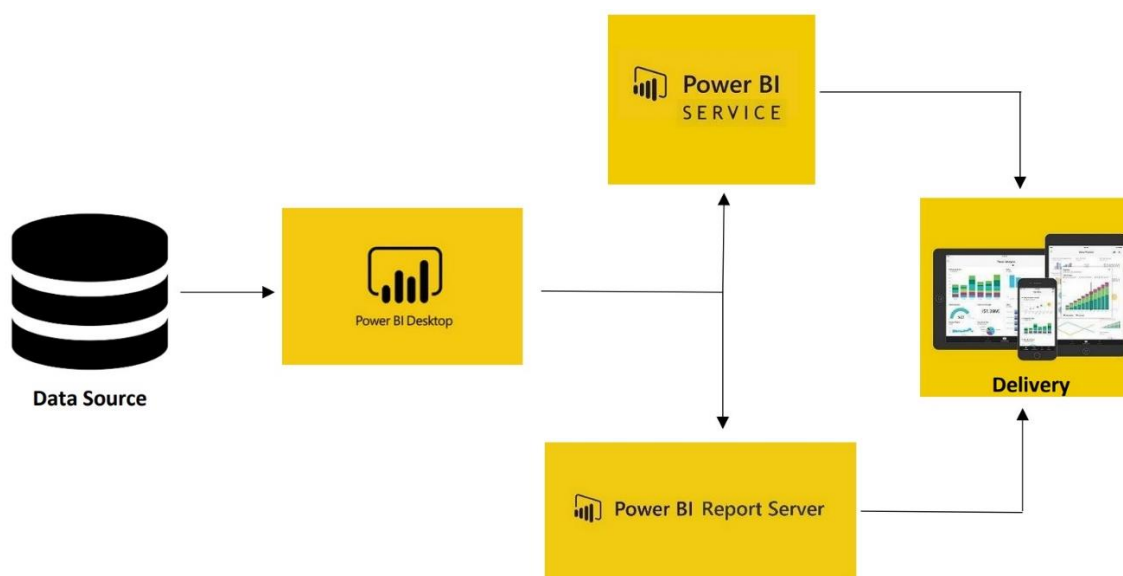
4.2 Components of Power BI Architecture

1. Data Sources

An important component of Power BI is its vast range of data sources. You can import data from files in your system, cloud-based online data sources or connect directly to live connections. If you import from data on-premise or online services there is a limit of 1 GB. Some commonly used data sources in Power BI are:

- Excel
- Text/CSV
- XML
- JSON
- Oracle Database
- IBM DB2 Database
- MySQL Database
- PostgreSQL Database
- Sybase Database
- Teradata Database
- SAP HANA Database
- SAP Business Warehouse server
- Amazon Redshift
- Impala
- Google Big Query (Beta)
- Azure SQL Database
- Salesforce Reports
- Google Analytics
- Facebook
- GitHub

2. Power BI Desktop

Power BI Desktop is a client-side tool known as a companion development and authoring tool.

This desktop-based software is loaded with tools and functionalities to connect to data sources, transform data, data modelling and create reports.

3. Power BI Service

Power BI Service is a web-based platform from where you can share reports made on Power BI Desktop, collaborate with other users, and create dashboards.

It is available in three versions:

- Free version
- Pro version
- Premium version

4. Power BI Report Server

The Power BI Report Server is similar to the Power BI Service. The only difference between these two is that Power BI Report Server is an on-premise platform. It is used by organizations who do not want to publish their reports on the cloud and are concerned about the security of their data.

Power BI Report Server enables you to create dashboards and share your reports with other users following proper security protocols. To use this service, you need to have a Power BI Premium license.

5. Power BI Gateway

This component is used to connect and access on-premise data in secured networks. Power BI Gateways are generally used in organizations where data is kept in security and watch. Gateways help to extract out such data through secure channels to Power BI platforms for analysis and reporting.

6. Power BI Mobile

Power BI Mobile is a native Power BI application that runs on iOS, Android, and Windows mobile devices. For viewing reports and dashboards, these applications are used.

7. Power BI Embedded

Power BI Embedded is an On-premises service in Azure. It offers APIs for embedding the reports and dashboards into custom applications. Till now, we have been discussing major components of the Power BI, and now, we will talk about the remaining components of Power BI as well.

Here is the list of the remaining Power BI Components.

8. Power BI Query

Power Query is the data connectivity that enables the business users to access the data which is stored in multiple data sources and redesign it to satisfy their business requirements. Power Query offers custom connectors SDK so that third-party users can create their data connectors.

9. Power Maps

Power BI Query is used to display how the values vary in proportion across the region. It also shows differences with the shadings that range from dark to light. It offers a 3D geospatial Data Visualization Tool.

10. Power Pivot

Power Pivot is an element that stores the information in memory and allows highly compressed data storage and incredibly quick aggregation and calculation. It is also accessible as part of Excel and can be used within an Excel workbook to build a data model. Power Pivot can load information on its own, or Power Query can load information into it. It is highly comparable to the tabular model of SSAS (SQL Server Analysis Services), which is like a Power Pivot server-based variant.

11. Power View

Power View offers interactive visualization that enables a drag-and-drop interface for users to create visualizations quickly and effectively in their Excel workbooks (using the Power Pivot data model).

12. Power Q&A

Power Q & A is the feature of Power BI, and it enables you to explore your data in your own words. In other words, you can use natural language and ask a question to get an answer from your data.

4.3 Power BI architecture – Working

We hope that you have understood the individual components of Power BI, and now, you will learn how these components work together. You will have a clear understanding of the Power BI Architecture with the help of the below image.

Working of Power BI Architecture
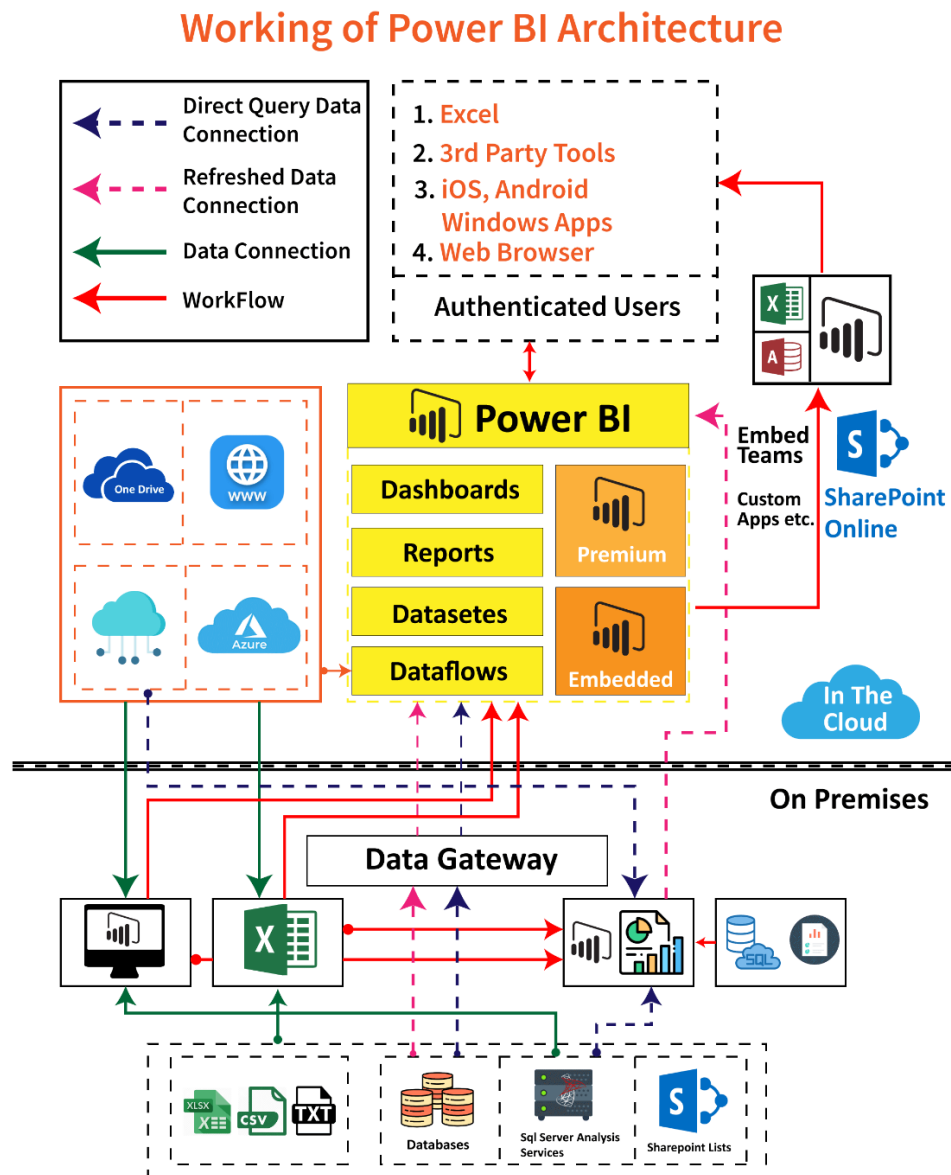
In the above diagram, it is clear that the upper half part represents On-Cloud services, and the lower half part represents the On-Premise services.

If you observe in the top of the image excel, web browsers and other sources are streaming into Power BI components, and they are called data sources. These data sources are authenticated users. Power BI has different data sources like On-Premise, Cloud databases, direct connections, etc.

On-Premise:

Power BI Desktop is accomplished with the authenticating, development and publishing tools. You can transfer the data from data sources to Power BI Desktop. And also, it allows users to create and publish reports on the Power BI Report Server or Power BI Service.

Power BI Publisher allows you to publish the Excel workbooks to the Power BI Report Server. Report Publisher and SQL server Data tools help in creating the KPIs, datasets, paginated reports, mobile reports, etc. All kinds of reports are published at the Power BI Report Server, and from there, reports are distributed to the end-users.

On-Cloud:

Power BI Gateway is the essential component in the Power BI architecture. The Power BI Gateway acts as a bridge or secure channel to transfer the data from On-premise data to On-cloud data sources or apps.

Cloud side architecture consists of a lot of components including Power suite having datasets, dashboards, reports, Power BI Premium, Power BI Embedded, etc. Users can embed the dashboards, reports into applications, SharePoint, Teams, etc. There are Cloud data sources and they are connected to the Power BI tools.

Power BI Service Architecture

In the previous section, you have learned how to publish the created reports in the Power BI Service.

Power BI Service enables the users to create and access the reports, dashboards from the client platforms like mobile devices, websites, etc. User needs to interact with the Power BI Service whenever they want to access the data that is created on the Power BI. So, now, we will learn how the Power BI Service works.

Power BI Service Architecture consists of two clusters. The following are the two clusters.

- Front End Cluster

- Back End Cluster

Now, we will discuss the two clusters in detail.

1. Front End Cluster: Front end cluster acts as an intermediate between the back-end cluster and the clients. It is also called a Web Front End Cluster. It establishes the initial connection and authenticates the users or clients using the Azure Active Directory. After user authentication, Azure Traffic Manager directs the user requests to the nearest data centres and Azure Content Delivery Network (CDN) allocates the statice files/content to the users or clients based on the geographical locations.

Front End
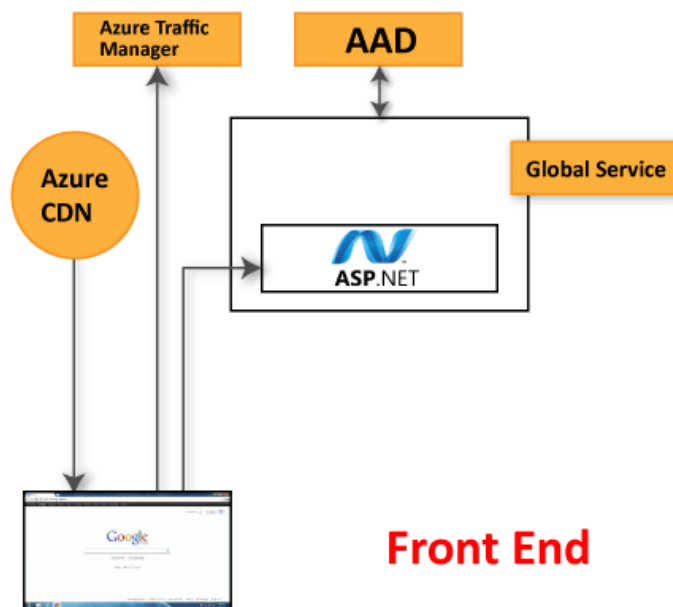
2. Back End Cluster: It manages the datasets, reports, storage, visualizations, data refreshing, data connections, and other services in the Power BI. At the back end cluster, the web client has only two direct points to interact with the data, i.e., Gateway Role and Azure API Management. These two components are responsible for authorizing, load balancing, routing, authentication, etc.
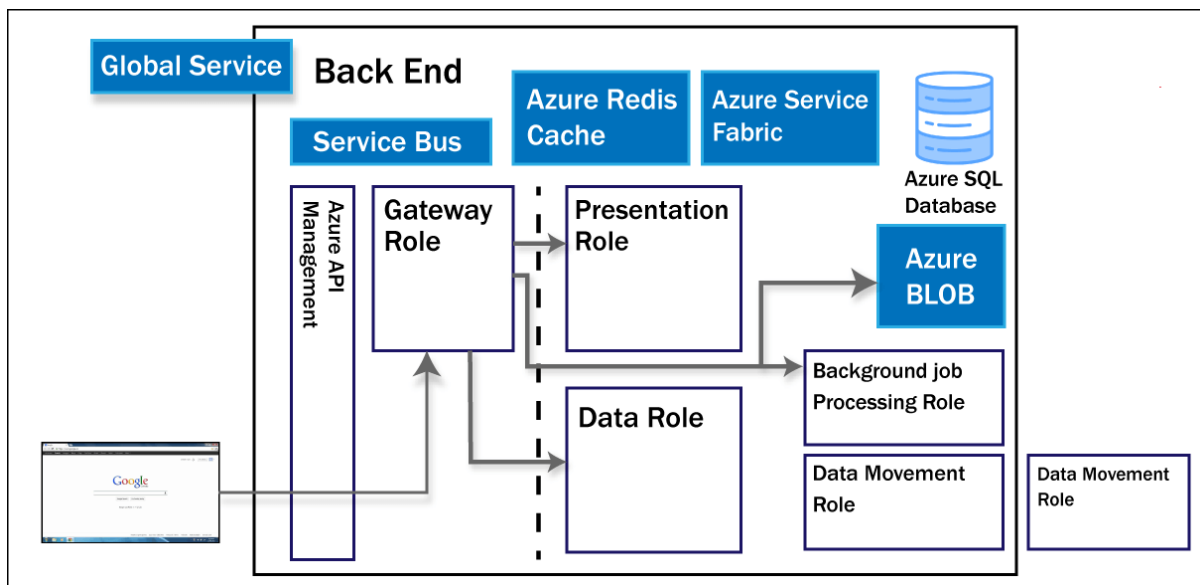
Working Of Power BI Service

- Power BI stores the data in two leading repositories, i.e., Azure SQL Database and Azure Block Storage. Azure Block Storage enables the users to store the datasets, and all system-related data and metadata are stored in the Azure SQL database.

- It authenticates the user requests and sends them to the Gateway Role. It processes the requests and assigns them to the appropriate components like Background Job Processing Role, Data Movement Role, Presentation Role, and Data Role.

- The presentation role manages all the associated visualization queries like reports and dashboards.

- Presentation Role sends requests to the Gateway Role to the Data Movement Role or Data Role for all relevant datasets.

- Azure Service Bus is used to connect and fetch the data from the On-Premises data sources with the cloud. It sends a request to execute the queries On-Premises data source and retrieve the data from its cloud service.

- The Azure Service Fabric allows all components and microservices which are related to the Power BI Service.

- Azure Cache helps in reporting the data that is stored in the in-memory of the Power BI system.

5. Deployment

5.1 Power BI Deployment

The deployment process lets you clone content from one stage in the pipeline to another, typically from development to test, and from test to production. During deployment, Power BI copies the content from the current stage, into the target one. The connections between the copied items are kept during the copy process. Power BI also applies the configured deployment rules to the updated content in the target stage. Deploying content may take a while, depending on the number of items being deployed. During this time, you can navigate to other pages in the Power BI portal, but you cannot use the content in the target stage.

5.2 Publish datasets and reports from Power BI Desktop

When you publish a Power BI Desktop file to the Power BI service, you publish the data in the model to your Power BI workspace. The same is true for any reports you created in Report view. You'll see a new dataset with the same name, and any reports in your Workspace navigator.

Publishing from Power BI Desktop has the same effect as using Get Data in Power BI to connect to and upload a Power BI Desktop file.
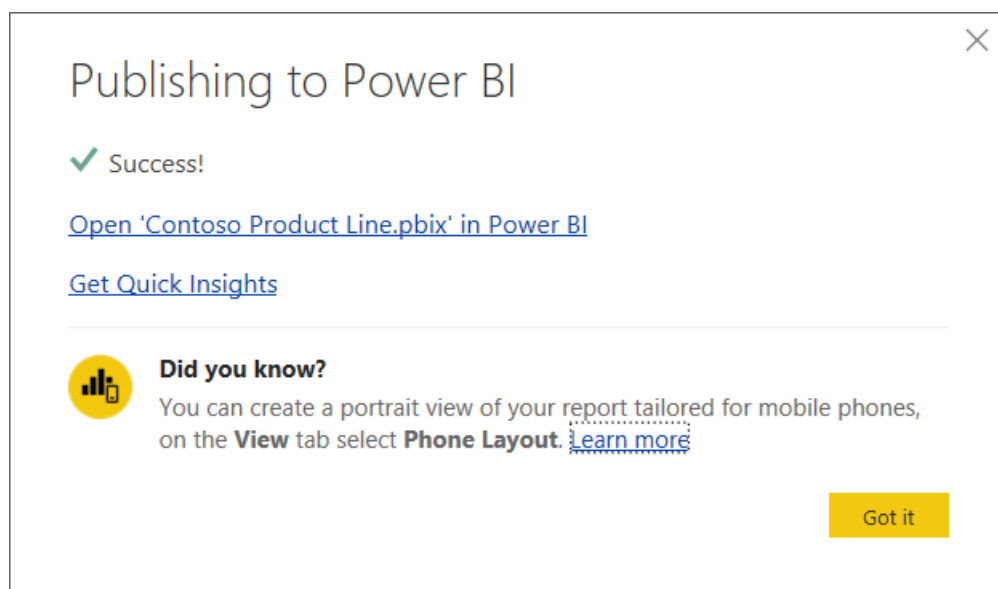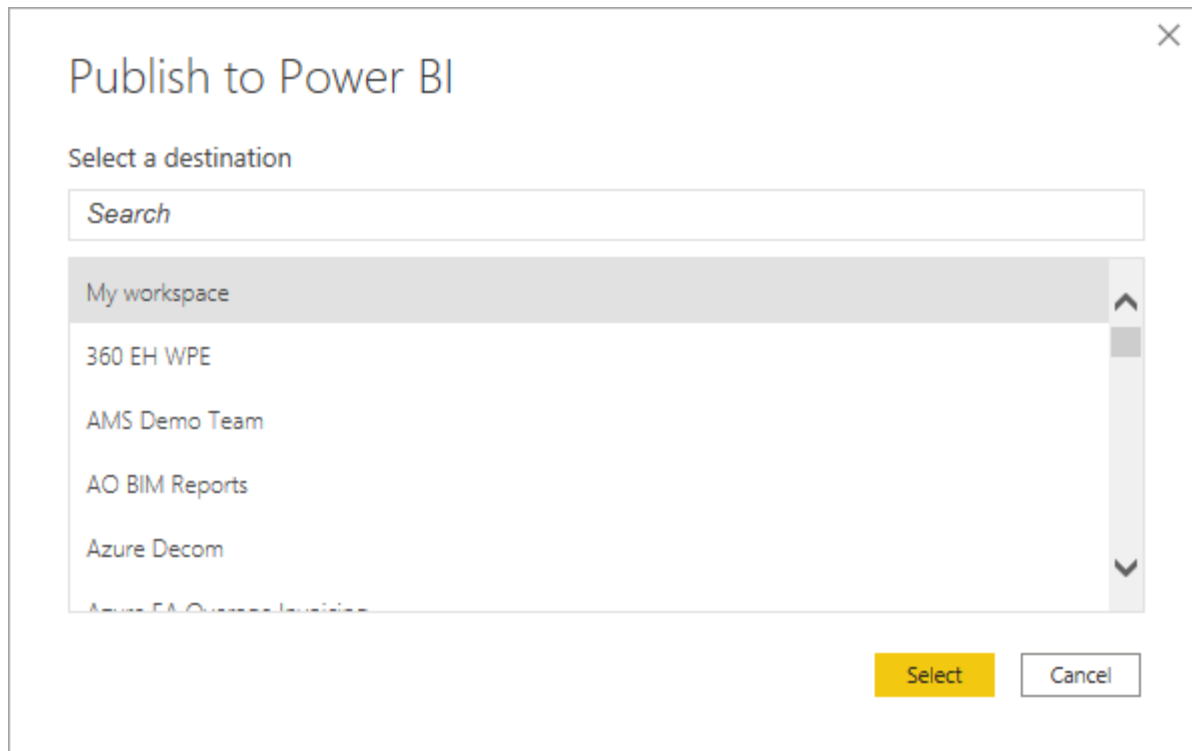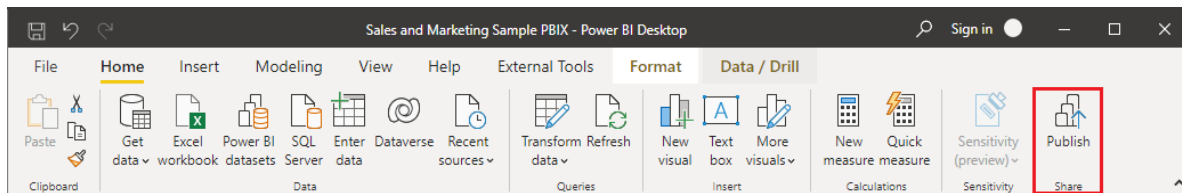
I created a Power BI Dashboard

US by number of bird strikes

Average of Altitude and Bird strikes count by Month and Phase

Phase ● Approach ● Climb ● Descent ● Landing Roll ● Parked ● Take-off run ● Taxi ● Bird strikes c...

937
772
1233
1828
2318
2078
3278
3710
3417
3038
1801
1019

January February March April May June July August September October November December

Month

Birdstrikes and Average of Altitude by Month and Precipitation

Phase ● Fog ● Fog, Rain ● Fog, Rain, Snow ● Fog, Snow ● Rain ● Rain, Snow ● Snow

Pie chart | bar chart

Month			
January	44.68%	39.01%	12.77%
February	24.24%	55.56%	14.14%
March	16.38%	67.24%	8.62%
April	13.74%	74.05%	
May	20.19%	76.53%	
June	19.17%	77.50%	

0% — Birdstrikes — 100%

avg. altitude | number of bird strikes | cost

Bird Strikes by Is Aircraft Large?, Indicated Damage

Pilot warned? ● N ● Y

Indicated Damage		
No – No damage	8.4K	6.8K	15.18K
Yes – No damage	4.6K	3.2K	7.80K
No – Caused damage	1.2K	1.85K
Yes – Caused damage	0.60K	

0K — 5K — 10K — 15K — Bird Strikes

Avg Altitude by Pilot warned?, Impact to flight

Is Aircraft Large? ● No ● Yes

2367.81
1757.85
2240.17
846.11
254.55
591.56
836.55
746.86
995.76
1470.29
831.43
638.86
985.09
633.65
351.44
1215.72
496.74
718.98
1605.83
853.47
752.35
0.35
0.39

Aborted Take-off | Engine Shut Down | None | Other | Precau... Landing | Aborted Take-off | Engine Shut Down | None | Other | Precau... Landing

N | Y
Impact to flight

Avg Cost by Year and Pilot warned?

Pilot warned? ● N ● Y

23K
12K
9K
11K
4K 5K 5K 5K 4K
4K 5K 4K 4K 7K 2K 3K 3K 6K

2000 2002 2004 2006 2008 2010
Year

US bird Strikes

Wildlife Struck by Remains of wildlife collected? and Remains of wildlife sent to Smithsonian

Remains of wildlife sen... ● False ● True

Remains of wildlife collected?		
False	45K	45K
True	17K	7K	24K

0K — 10K — 20K — 30K — 40K
Sum of Wildlife: Number Struck Actual

%GT Bird Strikes by Wildlife: Size and Wildlife: Species

Small Unknown bird - small 41%
Medium Unknow... 17%
Small Mourning d...
Small Eu... 2%
Large U... 3%
Small R...
Small Ba...
M...
M...
Sm...
Sm... 1% 1%
M...

%GT Cost by Wildlife: Size and Wildlife: Species

Large Canada goose 15%
Large White-tailed deer 12%
Medium Unk... 7%
Large Bald ea... 5%
Large Unkno... 5%
Large Turk... 5%
Small Whit... 4%
Large Grea...
Small Unk...
Small Eur... 2%
Large ...
Medium ...
Medium
Medium
Large ...
Mediu...
Me... 2%
Sm...
Me... 2%
1% 1%
S...