# TITANIC.

# TITANIC.

The hands-on stage - being able to work with data in practical ways.

**Key Abilities:**

Acquire raw data ad transform it into a tidy structure structured form, ready for analysis.
Use basic tools like Excel for analysing data and basic coding.
Formulate questions that clarify what you want to learn and drive meaningful analysis.
Interpret charts that use various techniques and understand the benefits of design simplicity.
Understand the basic statistics (averages, trends, distributions).



**Typical Learner Mindset:**

***"I can work with data, explore it, and draw meaningful conclusions"***

# TITANIC.

What questions can and (can not) be answered with data?

And once you are staring at some rows and columns, how do you start to make sense of the data in front of you?

In this activity you learn to take the first step in the data analysis process by asking good questions.

# AGENDA.

- Overview
- Introduction
- Manifest File
- Data Dictionary
- Initial Exploratory Questions
- Break
- Thoughts
- Deeper Questions
- Thoughts
- More Information Gathering
- Context Exploration
- Mr Andersson
- Conclusion

# OVERVIEW – 5 MINUTES.

- Increasingly, data arrives on our doorstep in the form of a **spreadsheet**.
  - Somebody sends you an Excel file or you download something from the web.
  - In this activity, you will work with colleagues to ask **good questions** of a spreadsheet so that you can take it from **data points** to **data story**.

- We will be using the **Titanic** data set which is based on the The Titanic - Machine Learning from a Disaster challenge which uses **machine learning** to create a model that predicts which passengers survived the Titanic shipwreck.

- For our purpose we will just be using the dataset which is attached.
  - There is some **important information** we need from the Kaggle Challenge that explains what the data is - or more exactly a **Data Dictionary** – we well get this this in a bit.

# OVERVIEW.

- **Participants**
  - Group of Twenty people.
  - Five Groups of Four people.

- **Materials**
  - A laptop for each group – along with a charger.
  - A copy of the Manifest.csv file (emailed).
  - A copy of the Manifest Modified.csv file (emailed).
  - Some basic familiarity with Excel.
  - Access to CoPilot.
  - Copy of questions.
  - Paper / Markers.

# INTRODUCTION – 2 MINUTES.

- The purpose of the workshop today is to **explore data**, **ask questions** and **derive insights**.

- The sinking of the Titanic is one of the most infamous shipwrecks in history.
- On April 15, 1912, during her maiden voyage, the widely considered “unsinkable” RMS Titanic sank after colliding with an iceberg.

- We have a copy of the Manifest (passengers) file from the Titanic that day.
  - Our first task will be to have a look the Manifest file together.

# MANIFEST FILE – 2 MINUTES.

| PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0 | 3 | Braund, Mr. Owen Harris | male | 22 | 1 | 0 | A/5 21171 | 7.25 | | S |
| **2** | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Thayer) | female | 38 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| **3** | 1 | 3 | Heikkinen, Miss. Laina | female | 26 | 0 | 0 | STON/O2. 3101282 | 7.925 | | S |
| **4** | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35 | 1 | 0 | 113803 | 53.1 | C123 | S |
| **5** | 0 | 3 | Allen, Mr. William Henry | male | 35 | 0 | 0 | 373450 | 8.05 | | S |
| **6** | 0 | 3 | Moran, Mr. James | male | | 0 | 0 | 330877 | 8.4583 | | Q |
| **7** | 0 | 1 | McCarthy, Mr. Timothy J | male | 54 | 0 | 0 | 17463 | 51.8625 | E46 | S |
| **8** | 0 | 3 | Palsson, Master. Gosta Leonard | male | 2 | 3 | 1 | 349909 | 21.075 | | S |

# DATA DICTIONARY – 2 MINUTES.

| Variable | Definition | Key |
|---|---|---|
| **Survival** | Survival | 0 = No, 1 = Yes |
| **Pclass** | Ticket class | 1 = 1st, 2 = 2nd, 3 = 3rd |
| **Sex** | Sex | |
| **Age** | Age in years | |
| **Sibsp** | # of siblings / spouses aboard the Titanic | |
| **Parch** | # of parents / children aboard the Titanic | |
| **Ticket** | Ticket number | |
| **Fare** | Passenger fare | |
| **Cabin** | Cabin number | |
| **Embarked** | Port of Embarkation | C = Cherbourg, Q = Queenstown, S = Southampton |

**Pclass**: A proxy for socio-economic status (SES)
1st = Upper
2nd = Middle
3rd = Lower

**Age**: Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5

**Sibsp**: The dataset defines family relations in this way...
Sibling = brother, sister, stepbrother, stepsister
Spouse = husband, wife (mistresses and fiancés were ignored)

**Parch**: The dataset defines family relations in this way...
Parent = mother, father
Child = daughter, son, stepdaughter, stepson
Some children travelled only with a nanny, therefore Parch=0 for them.

# INITIAL EXPLORATORY QUESTIONS – 25 MINUTES.

- Email the Manifest file and distribute the **Initial Exploratory Question Set**.
- In groups of 5 (4 stations).

- Try getting to know the data.
- I have distributed a set of questions.
  - Feel free to have a try at any of them and we will go through the answers after.

- **Don't feel you have to answer** the questions though.

- Please do not use any other tabs – they are for later activities.

# INITIAL EXPLORATORY QUESTIONS.

- **Getting to know the Data**

  - **What are** the different columns (features) available in the dataset?
  - **How many** passengers are listed in the manifest? **891**
  - **What does** the "Survived" column represent?
  - **How many** passengers survived versus did not survive? **342** Survived / **549** Perished.

# INITIAL EXPLORATORY QUESTIONS.

- **Basic Demographics**

  - **What are** the different passenger classes ("Pclass")?
  - **How many** passengers are there in each class? (**1st – 216, 2nd – 184, 3rd – 491**)
  - What is the gender distribution?
    - **How many** males and females are on board? (**Male 577, Female 314**)
  - **What is** the age range of passengers? (**Youngest about 5 months, Oldest 80**)
  - **What is** the average age? (**Average Age 29.7**)

# INITIAL EXPLORATORY QUESTIONS.

- **Family and Travel Companions**

  - What do the columns "SibSp" and "Parch" represent?
    - **"SibSp"** = Number of siblings/spouses aboard
    - **"Parch"** = Number of parents/children aboard
  - **How many** passengers travelled alone (SibSp = 0 and Parch = 0)? **About 537**
  - Can you find the passenger(s) with the largest family onboard? **8 family members**

# INITIAL EXPLORATORY QUESTIONS.

- **Ticket and Fare**

  - **What is** the range of ticket prices ("Fare")? (**$0 to $512**)
    - **What is** the average fare? (**$32.0**)
  - Is there a correlation between passenger class ("Pclass") and fare? **Yes**

# INITIAL EXPLORATORY QUESTIONS.

- **Embarkation Points**

  - **What are** the different embarkation points ("Embarked")? (**Southampton, Cherbourg, Queenstown**)
  - **Which** embarkation point has the most passengers? (**Southampton 644**)

# INITIAL EXPLORATORY QUESTIONS.

- **Survival Analysis**

  - **What percentage** of passengers survived? (**38%**)
  - **Compare survival rates** between **males** (**19%**) and **females** (**74%**)

# BREAK – 10 MINUTES.

# THOUGHTS – 5 MINUTES.

- The Initial Exploratory Question set is interesting for pulling out specific data points, but they are not really showing any **deeper correlations** between the data.

- Lots of **How Many** type questions.

- How did you get on with the final **compare** question?
  - With this question we have to use a bit more **logic** as we are asking within the class male/female what are the gender survival rates.
  - This is far different than **just quoting** the numbers.
  - Does this new logic show **more of an insight**?

# DEEPER QUESTIONS – 10 MINUTES.

- Distribute the Manifest Modified.csv file to participants.

- Specifically, only look at the following Tabs:

  - **Survival_By_Class**
  - **Survival_By_Gender**
  - **Survival_By_Age**

- In your groups look at the tabs listed above.

# DEEPER QUESTIONS.

- **Survival_By_Class**
  - Using the data in the "**Survival_By_Class**" tab, **analyse** and **compare** the survival rates of passengers across the three different classes (**First**, **Second**, and **Third** Class).
  - What insights can you draw about how passenger class affected the likelihood of survival during the Titanic disaster?

- **Survival_By_Gender**
  - Using the "**Survival_By_Gender**" tab, **compare** the survival rates between male and female passengers.
  - What does the data reveal about the **differences** in survival likelihood based on gender?
  - Discuss possible **social** or **historical** factors that might explain these **differences**.

- **Survival_By_Age**
  - Using the **Survival_By_Age** tab, examine the **survival rates** across different age groups.
  - Which age group had the **highest** survival rate, and which had the **lowest**?
  - What **trends** or **patterns** do you observe regarding age and survival, what factors could have contributed to these?

# THOUGHTS – 10 MINUTES.

- **Survival_By_Class**
  - What is interesting here is that the datapoints for the **actual** number of survivors do not reveal that much.
  - It's only when expressed as a **percentage of overall survivors for that class** that you see that first class had a much more significant chance of survival.

- **Survival_By_Gender**
  - Again, just listing the number of male and female survivors **doesn't give you much insight**.
  - By looking at the **percentage** of only male survivors and the percentage of only female survivors, reveals a far larger gap in the likelihood of survival.

- **Survival_By_Age**
  - This question guides us to analyse survival data by age categories, encouraging us to identify patterns and think about **possible reasons** behind the age-related survival differences.

# THOUGHTS.

- There were a lot more **Compare** type questions.
- Now we have started to **analyse** the data and look in to more of the context and **form ideas or hypothesis** about the data.
- Did you find the **data lacking**, and there were **more questions** you found you wanted to ask.
- Did you find looking into the context revealed **other insights**.

# MORE INFORMATION GATHERING – 25 MINUTES.

- Now we will look at other accompanying information:
  - **Fare Tab**
  - **Embarked Tab**
  - **Parch Tab (Parents/Children Abord)**
  - **SibSp Tab (Siblings/Spouses Abord)**
  - **Cabin Tab**

- Each Group to take a tab each and investigate.
  - **Group 1** – Fare Tab
  - **Group 2** – Embarked Tab
  - **Group 3** – Parch Tab / SibSp Tab (… these two tabs are closely related)
  - **Group 4** – Cabin Tab

- After 5 minutes the facilitator will give you a help.

# MORE INFORMATION GATHERING.

- **Group 1 - Fare Tab**

  - Using the Fare tab, analyse how the survival rates vary among passengers who paid different fare amounts.
    - Which fare **range** had the **highest** survival rate?
    - How might the fare paid relate to other factors such as **passenger class** or **socio-economic status**?

# MORE INFORMATION GATHERING.

- **Group 2 - Embarked Tab**

  - Refer to the Embarked tab to compare survival rates of passengers based on their port of embarkation (**Cherbourg**, **Southampton**, **Queenstown**).
  - Which embarkation point had the **highest** survival rate?
    - What reasons might explain the **differences** in survival rates among these groups?

# MORE INFORMATION GATHERING.

- **Group 3 - Parch Tab (Parents/Children Abord)**

  - Using the Parch tab, investigate the **survival rates** of passengers traveling with **different** numbers of parents or children aboard.
    - How does traveling alone versus with family members **affect** survival chances?
    - What might this tell us about the **role** of family presence during the disaster?

# MORE INFORMATION GATHERING.

- **Group 3 - SibSp Tab (Siblings/Spouses Aboard)**

  - Examine the **SibSp tab** to determine how the number of siblings or spouses aboard **influenced** passenger survival.
    - Are passengers traveling alone **more or less** likely to survive compared to those with one or more siblings/spouses?
    - Discuss potential reasons for this pattern.

# MORE INFORMATION GATHERING.

- **Group 4 - Cabin Tab**

  - Using the **Cabin tab**, **compare** the survival rates of passengers with a recorded cabin number versus those without one.
    - What **differences** do you observe?
    - How might having a cabin number (and thus likely a cabin location or class) have **impacted** a passenger's chance of survival?

# MORE INFORMATION GATHERING.

- **Feedback**

  - **Group 1** – Fare Tab
  - **Group 2** – Embarked Tab
  - **Group 3** – Parch Tab / SibSp Tab
  - **Group 4** – Cabin Tab

# CONTEXT EXPLORATION – 15 MINUTES.

- **Instructions**
  - Have a look at the following tabs and what they represent:
    - **Group 1 - Survival_By_Age_Gender**
    - **Group 2 - Survival_By_Fare_Class**
    - **Group 3 - Survival_by_Embarkation_Class**
    - **Group 4 - Survival_by_Cabin_Class**
  - Then use CoPilot or any other tool to see if there is any context or maritime law around why some of these figures are giving this information.

# CONTEXT EXPLORATION.

- **Survival_By_Age_Gender**

  - The widely practiced maritime custom of "**women and children first**" (also known as the Birkenhead Drill) influenced evacuation priorities during the Titanic disaster.
    - This code of conduct prioritised saving women and children before men, which explains **higher survival rates** among females and younger passengers.

  - This practice was informal and not legally binding but was deeply ingrained in maritime **tradition**, reflecting the social norms of the time.

# CONTEXT EXPLORATION.

- **Survival_By_Fare_Class**

  - Passenger class on ships like the Titanic **reflected** social hierarchy and economic status, with first-class passengers often having better access to lifeboats and safer, **more accessible** locations on the ship.
  - Maritime regulations before the Titanic disaster did not require enough lifeboats for all passengers, but lifeboats were often allocated based on class or status.
    - This **inequality** contributed to higher survival rates among wealthier passengers who could afford more expensive tickets.

  - Post-Titanic, maritime law was reformed (e.g., SOLAS 1914) to require **sufficient** lifeboats for all passengers regardless of class.

# CONTEXT EXPLORATION.

- **Survival_by_Embarkation_Class**

  - Embarkation points (ports of boarding) often **reflected** passenger demographics and socioeconomic status.
    - For example, Southampton was the **main** boarding port for many third-class passengers (often immigrants or working class), while Cherbourg and Queenstown had more first- and second-class passengers boarding.

  - Lower survival rates for certain embarkation points **correlate** with where more third-class passengers boarded, who faced physical and social barriers (such as **restricted** access to upper decks and lifeboats) due to ship design and policies.

  - Maritime safety laws at the time **did not adequately** address these disparities, which were later scrutinized and addressed in safety reforms.

# CONTEXT EXPLORATION.

- **Survival_by_Cabin_Class**
  - Cabin assignment often **correlated** with passenger class and location on the ship.
    - First-class cabins were typically situated on upper decks closer to lifeboats, providing **easier** and **quicker** access during evacuation.
  - Third-class cabins were generally located in lower decks and more **isolated** parts of the ship, sometimes behind locked gates or barriers, **limiting** passengers' ability to reach lifeboats quickly.
  - These design and operational factors **contributed** to survival disparities between passengers with cabin numbers (mostly first/second class) and those without (often third class or steerage).
  - After the Titanic, maritime regulations were improved to ensure **better access** to lifeboats for all passengers, regardless of cabin location.

# MR ANDERSSON – 10 MINUTES.

- What story can we tell about **Mr Andersson**?

- Does the data make **sense**?

  - **How** did family size and composition impact survival chances?
  - **What** does the Andersson family tell us about the experiences of third-class passengers?
  - **How** might the family have coordinated or struggled during evacuation?
  - **What** broader social or structural factors affected families like the Anderssons?

# MR ANDERSSON.

- **How did family size and composition impact survival chances?**

  - Larger families, like the Anderssons, faced **significant** challenges during evacuation due to coordinating multiple members, especially with young children.

  - Survival chances could be **lower** for large families in lower classes because it was **harder** to keep everyone together and access lifeboats quickly.

  - However, children and women were **prioritised**, which may have **increased** survival chances for the younger and female family members compared to adult males.

  - Family size could both be a support system and a logistical obstacle in a crisis.

# MR ANDERSSON.

- **What does the Andersson family tell us about the experiences of third-class passengers?**

  - The Anderssons were a third-class family, **reflecting** the immigrant or working-class demographic prevalent in that section.

  - Their experience **illustrates** the difficulties faced by third-class passengers: less access to upper decks, potential language **barriers**, and physical separation from lifeboats.

  - Despite these barriers, some family members survived, showing that survival was possible but often depended on quick access and situational factors.

  - The family's story **highlights** the social inequalities onboard that affected who could escape.

# MR ANDERSSON.

- **How might the family have coordinated or struggled during evacuation?**

  - **Coordinating** a large family during the chaos would have been extremely **difficult,** especially with young children and limited information.

  - Physical barriers, locked gates, and crowding in third class likely **complicated** their movement to lifeboats.

  - Parents might have **prioritised** children's safety over their own, consistent with "women and children first" practices.

  - Language and unfamiliarity with the ship could have added to **confusion** and **delays**.

# MR ANDERSSON.

- **What broader social or structural factors affected families like the Anderssons?**

  - The ship's class segregation physically **separated** third-class passengers from lifeboat access points and crew assistance.

  - Maritime safety regulations at the time did not mandate enough lifeboats for all, **disproportionately affecting** lower-class passengers.

  - Social norms and prejudices likely **influenced** how crew and other passengers prioritised evacuation.

  - Post-Titanic reforms aimed to address these **inequalities** by improving lifeboat capacity and access for all passengers regardless of class.

# MR ANDERSSON.

- **August Edvard Andersson**

  - He was not a member of the Andersson family from Kättestorp that perished. He was a **survivor** who was traveling under an assumed name to avoid legal trouble.

  - The Alias: He was a Swedish journalist and political activist who had written articles critical of the Swedish King.
    - To escape prosecution, he travelled under the pseudonym **August Wennerström**.

# CONCLUSION.

- How could you **improve** the visuals?
  - One page summary?

- Did anyone notice this was **not** a complete manifest?
  - It represents only 50% of the passengers
  - Remember to always **check your data** …