

KTDK: Kuch Toh Data Kahenge

About The Event:

This is an individual participation event. In the event, the participants will be provided with excel sheets containing some data (number of sheets will be disclosed on the day of the event itself), which would consist of various columns and a special answer column which would uniquely classify the rows of the data into categories. And the category(answer column) of the row would directly or indirectly depend to some or more extent on each of the columns in the data. And the participants would have to provide their guesses about the relation of each column on the special answer column to have the best prediction possible.

Datasets:

The excel spreadsheet containing data would be referred to as the dataset. The columns and rows will be referred to as features and tuples respectively.

The features(column) would be of 2 kinds: Categorical Features and Value-Based Features.

Categorical Features would have discrete values(with description) while the value-based features would have continuous value in a range.

Example: The Titanic Dataset, it contains the data of all the passengers who boarded the historic Titanic Ship. In this dataset, the features are as follow:

PID: Unique ID for each passenger.

PClass: The Class of ticket the passenger brought.

Name: Name of passenger.

Sex: The gender of passenger.

Age: Age of Passenger

SibSp: The number of spouses and/or siblings of the passenger on board.

ParCh: The number of parents and/or children of the passenger on board.

Ticket: Ticket No. of the passenger.

Fare: Charges of the ticket.

Cabin: Cabinet No. of the passenger.

Embarked: Port of Embarkment.

In this dataset, Survived is the answer, where it determines if the passenger survived(1) or not(0).

You can view the dataset at this link:

https://docs.google.com/spreadsheets/d/1iD4yZ-1IakWTPFE5E0tM_XxwbYPCblxfQ8OwPvmh714/

This is just a sample dataset and it will not be used in the event.

In this dataset, the features sex, Embarked and PClass are categorical features as these features suggest the following:

Sex: if 1 the passenger was male, and if 0 was female.

PClass: if the PClass is 1, 2, 3 then the passenger had upper, middle, lower class respectively.

Embarked: if the value is 1, 2, 3 then the passenger's port of embarkment was Queenstown, Cherbourg, Southampton respectively.

While other features like Fare, ParCh and SibSp can have any integer or float value for them. And the other features can have a string value.

Guessing the relation:

Participants have to assign weights(A number between 0 and 1 inclusive) to each and every feature, which would define to what extent the feature affects the answer.

Weights for categorical features would be binary i.e. either 0 or 1. If the participants think a feature has no effect on the prediction then the weight would be 0 and if you think it has an effect then it would be 1.

While for the value-based features the weight would be any value in the range of 0 to 1. If the effect is more, then the weight would tend to 1, else if it has less effect it would tend to 0 and if 0 is selected then the feature will be dropped.

Example:

In the above dataset, PClass(Passenger Class) feature would definitely affect the survival factor so we can assign its weight to 1 (As a Categorical feature, only 0 or 1 is allowed as weight).

Fare feature would have such a big impact on survival as if the fare is more the passenger was near the upper deck, so we can assign its weight to 0.8. (As for a Value-Based feature, anything between 0 and 1 inclusive is allowed as weight).

Tips:

The participants can use the charts and graphs tool available in Google Sheets and MS-Excel to find out value information from the dataset, by visualizing the data. Because numbers never lie.

For Visualisation of the data:

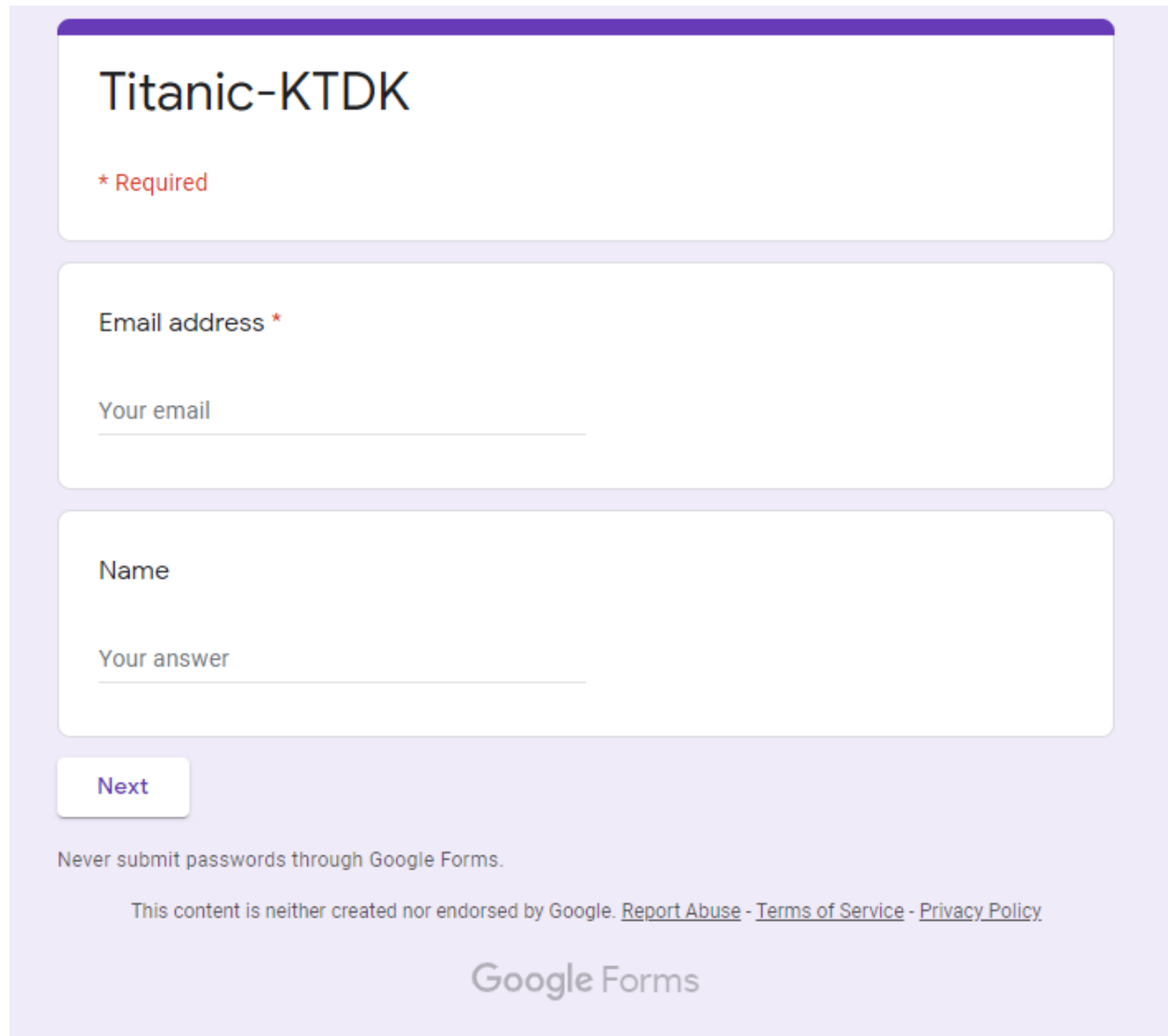
In Spreadsheet: <https://www.youtube.com/watch?v=IFX9e9SqGlw>

In MS-Excel: <https://www.youtube.com/watch?v=Z2gzLYaQatQ>

Submission:

Submission will be taken using Google Forms, and more details regarding the same will be informed on the day of the event.

Example:



The image shows a Google Form titled "Titanic-KTDK". Below the title, there is a red asterisk followed by the word "Required". The form contains two text input fields. The first field is labeled "Email address *" and has a placeholder text "Your email". The second field is labeled "Name" and has a placeholder text "Your answer". Below the second field, there is a "Next" button. At the bottom of the form, there is a warning: "Never submit passwords through Google Forms." followed by a line of text: "This content is neither created nor endorsed by Google. [Report Abuse](#) - [Terms of Service](#) - [Privacy Policy](#)." The Google Forms logo is at the very bottom.

Titanic-KTDK

* Required

Email address *

Your email

Name

Your answer

Next

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. [Report Abuse](#) - [Terms of Service](#) - [Privacy Policy](#).

Google Forms

Titanic-KTDK

Submission

PClass

☐ 0

☐ 1

Name

☐ 0

☐ 1

Sex

☐ 0

☐ 1

Age

Number between 0 and 1(inclusive)

Your answer

SibSp

Number between 0 and 1(inclusive)

Your answer

ParCh

Number between 0 and 1(inclusive)

Your answer

Ticket

☐ 0

☐ 1

Fare

Number between 0 and 1(inclusive)

Your answer

Cabin

☐ 0

☐ 1

Embarked

☐ 0

☐ 1

[Back](#)

[Submit](#)

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. [Report Abuse](#) - [Terms of Service](#) - [Privacy Policy](#)

Google Forms