Tutorial 3h - Side Inputs and Outputs Apache Beam:

Side Inputs:

- A side input is an additional input that your DoFn can access each time it processes an element in the input PCollection.
- In addition to the main input PCollection, you can provide additional inputs to a ParDo transform in the form of side inputs.
- Side inputs are useful if your ParDo needs to inject additional data when processing each element in the input PCollection, but the additional data needs to be determined at runtime (and not hard-coded).
- Side inputs must be small in size and not as big as pcollection because it has to be kept in memory of each worker
- Such values might be determined by the input data, or depend on a different branch of your pipeline.

Additional outputs:

- While ParDo always produces a main output PCollection (as the return value from apply), you can also have your ParDo produce any number of additional output PCollections.
- If you choose to have multiple outputs, your ParDo returns all of the output PCollections (including the main output) bundled together.

Resources:

- https://beam.apache.org/documentation/patterns/side-inputs/
- https://beam.apache.org/documentation/programming-guide/#additional-outputs
- https://beam.apache.org/documentation/programming-guide/#side-inputs