# Data Science Capstone – Project Description

## Introduction

Let’s imagine we are working at a tourist agency in Germany that offers services to foreign visitors. International tourists often ask themselves which cities in their destination country they should visit. Given their limited time and resources, we need to advise them on which cities are the most worthy ones of their attention.

So far, our intuitions and expert opinions have guided the advice we give. However, data-driven and evidence-based reasoning is gaining momentum among our target audiences. They are used to online websites showing them quantified information and rankings of cities, attractions, restaurants and many other things. We can see that we are losing ground to these competitors.

So we decide to set up our own data-driven model to advise our customers on which cities in Germany they should visit. In the long run, we want to implement a sophisticated system that can be used in various ways to support our work. But for now, we start by setting up a simple analysis that we can then refine over time.

## Data

We use a dataset of major German cities, i.e. cities with more than 100,000 inhabitants. Important datapoints per city that we use for our analysis are the geographic coordinates and the population size. For that, we download a comprehensive list of all cities worldwide from the following website: <https://simplemaps.com/data/world-cities>. We then load it into a dataframe and filter for German cities with more than 100,000 inhabitants.

We then leverage Foursquare location data to get a list of top recommended venues around the center of each city and see how many of these venues are highly relevant to tourists (e.g. monuments, historic sites). This allows us to calculate a “Tourism Score” that indicated tourist attractiveness for each city.