# Der Touristik Köln Operations Optimization

Der Touristik Köln (DTK hereafter) is a German travel company which leverages analytics to optimize its operations (https://www.sas.com/en_us/customers/der-touristik-de.html). They must purchase enough airplane tickets and hotel rooms to accommodate all their potential customers while not preparing for too many customers. Additionally, they must try to optimize their purchase timing to ensure that they are getting the best deal possible. They attain their goals using a series of analytical models and data warehouse solutions developed by SAS.

I imagine the system that DTK uses looks something like this:
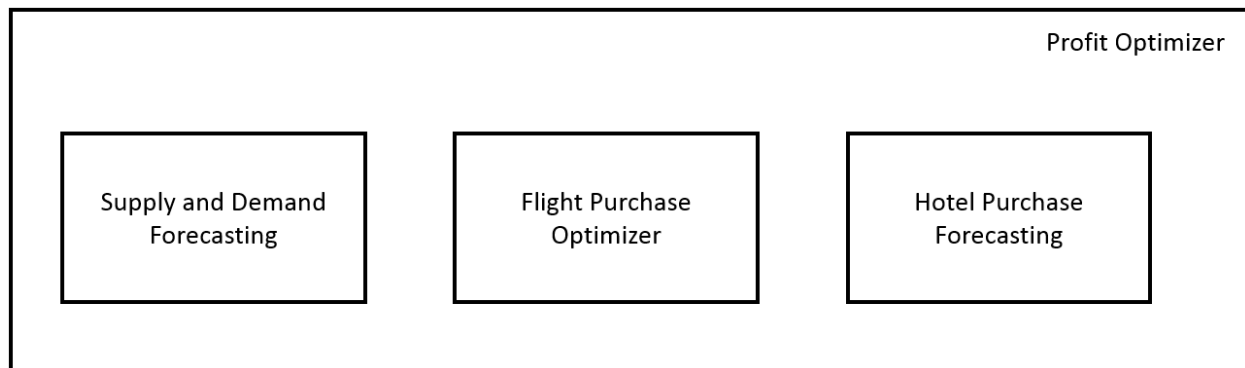
where **Supply and Demand Forecasting** is a model for determining the ideal price per customer and the number of customers to prepare for, and **Flight Purchase Optimizer** and **Hotel Purchase Forecasting** are models for predicting flight and hotel prices respectively. These models are used by the **Profit Optimizer** to maximize the profit of the company's operations.

Supply and Demand Forecasting

This model determines the optimum price per customer and the number of customers to purchase airfare and hotel for to maximize revenue (cost per customer x number of customers). The process starts with an A/B test which would be carried out with a special online advertising campaign, separate from DTK's standard, perpetual advertising methodology. For this A/B test we are trying to determine how the pricing of a travel package affects the amount of people who purchase it. The A/B test results may look something like this:

	Cost	% purchased (r_p)
A	$1.05\, c_m$	3%
B	$1.10\, c_m$	2.7%

where c_m is the minimum cost for the offered travel package – this is the break-even point for DTK. r_p can be converted to an expected number of customers, g, by

$$g = r_p E$$

where E is the exposure. This is the total amount of people exposed to DTK's products. This could be via DTK's perpetual advertising efforts, prior experience with DTK or word of mouth. The result of this A/B test would be used differently than described in the lectures for the course. Rather than choosing the better of A/B, we take advantage of the performance of these two scenarios to estimate the demand curve for the travel packages. The A/B test gives us two data points on the demand curve. If we assume a linear relationship we get the following plot:
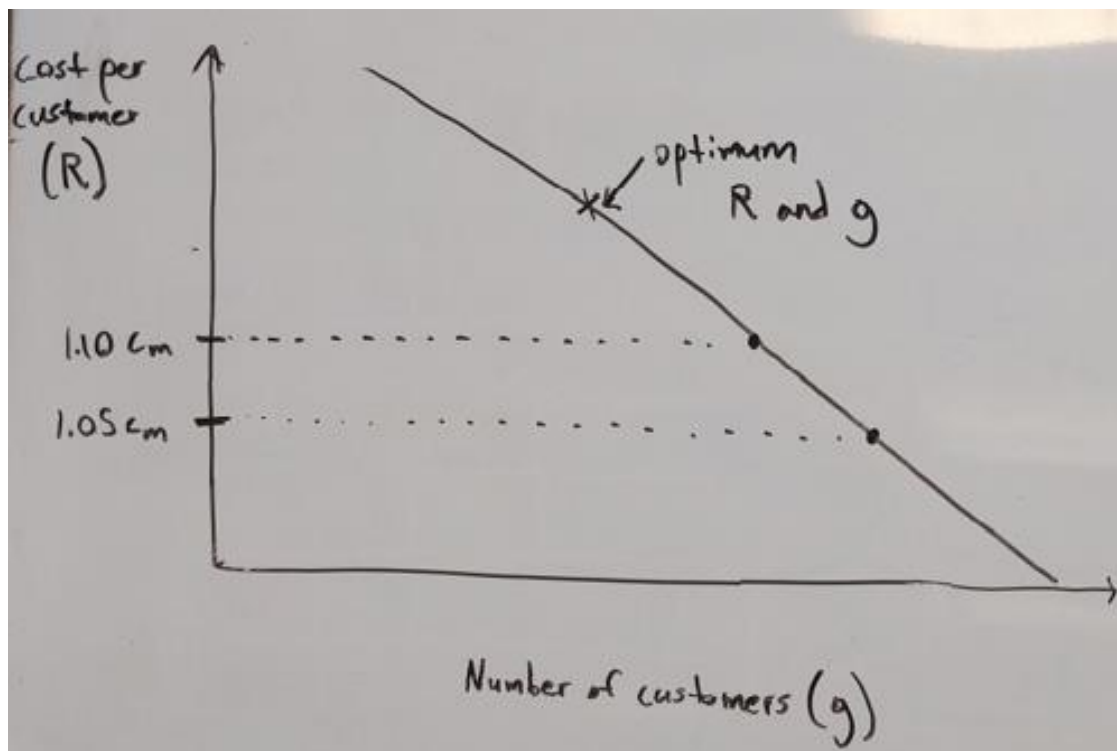
From this plot we can choose the optimum cost per customer (R) and number of customers to prepare for (g). This is the point which lies on the demand curve and maximizes the revenue (R x g).

Flight Purchase Optimizer

One of the primary drivers of cost for DTK is timing of airline ticket purchase. The optimum number of days ahead of the departure day to purchase airfare can depend on the route and the time of year. I imagine that they use an ARIMA model for this task. Specifically, I would expect a unique model for each origin/ destination/ number of days ahead of flight combination. This meta-model would then be used to determine the best purchase timing for combinations of destinations/origins/departure dates.

An example of what the historical data set for training the ARIMA models underlying the meta-model is shown below. This is for one unique combination of origin and destination (O_1 and D_1). The full meta-model would be composed of one of these for each unique combination of origin and destination.
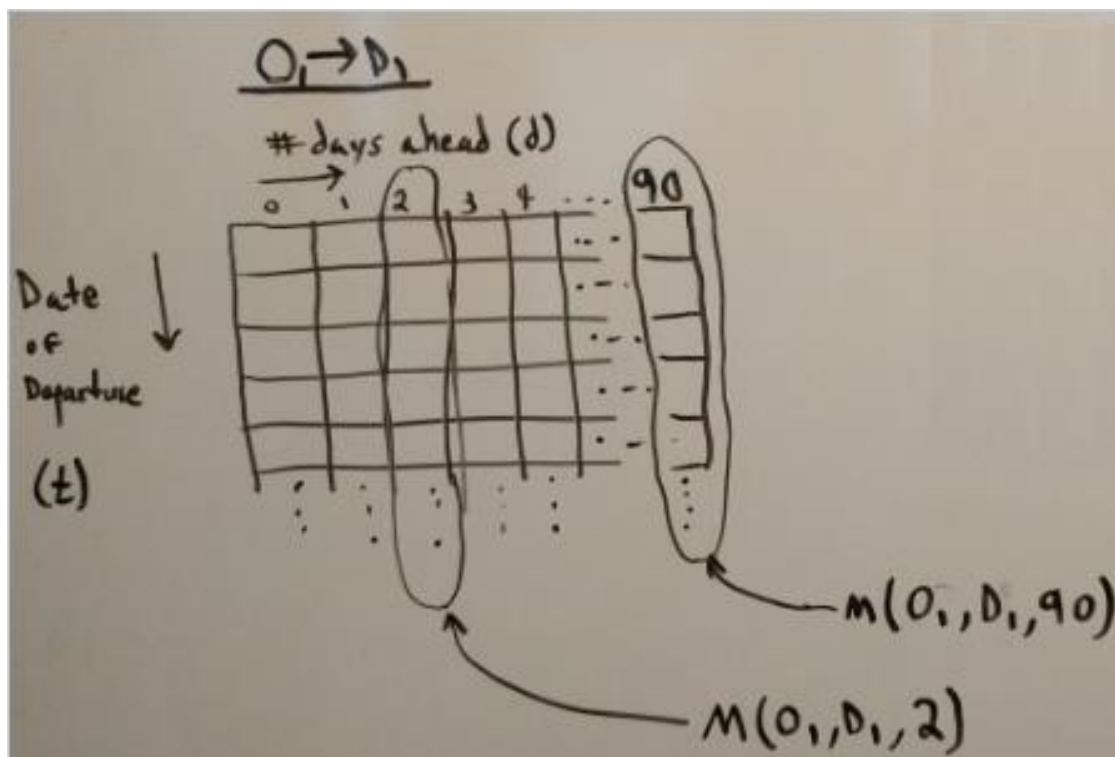
Each row is a different date of departure and each column is a different number of days ahead of time to purchase tickets. For this arrangement of data, we could have unique ARIMA models for each column. Here we introduce the notation M(O, D, d) to denote the ARIMA model corresponding to the Oth origin airport, Dth destination airport and d-day-ahead purchase

strategy. The two example models ($M(O_1, D_1, 2)$ and $M(O_1, D_1, 90)$) are shown on a timeseries plot below.
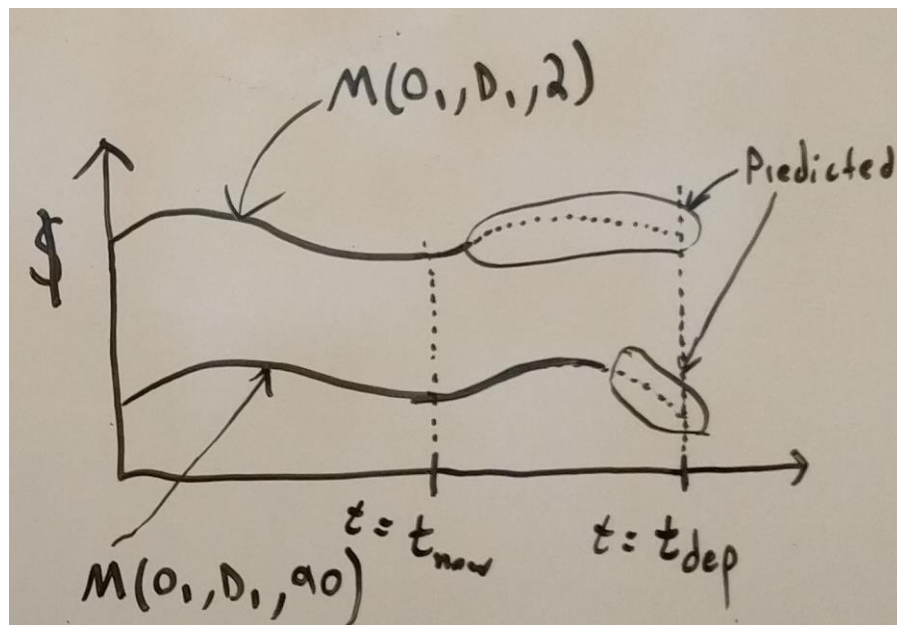
Here, t_{now} is the present time and t_{dep} is the desired departure date. Note that the price has seasonality and that the 2-day-ahead purchase strategy is more expensive than the 90-day-ahead strategy. The dotted portions of the plots represent dollar amounts forecasted by the ARIMA model.
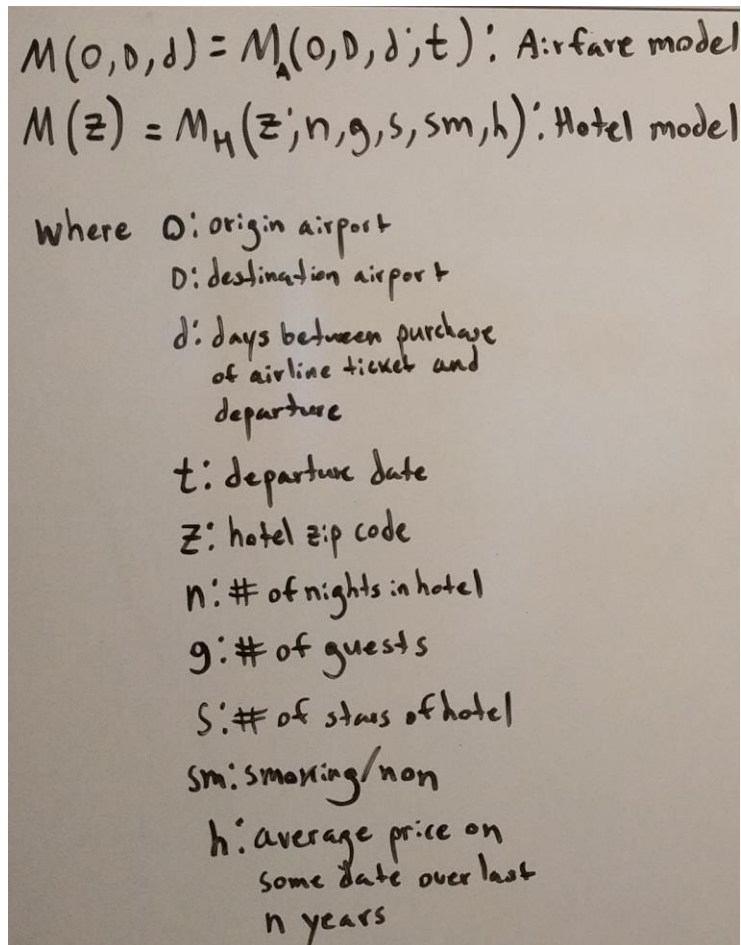
Hotel Purchase Forecasting

For the Hotel Purchase Optimizer, we will assume that DTK uses a regression-based meta-model with one regression model for each destination zip code that DTK serves. Each zip code model would leverage the following factors:

- Number of guests
    - More guests will increase the total cost, but the per guest cost may go down
- Number of nights of stay
    - More nights will increase total cost, but the per night cost may go down
- Number of stars of hotel
    - Some guests may want to stay in a luxury hotel. Need to know how this affects price
- Smoking or non-smoking
- Average price of similar room on same date over last n years
    - Determine similar room via KNN if necessary

We will denote the Hotel Purchase Optimizer for zip code Z as M(Z).

Profit Optimizer

The profit optimizer uses the aforementioned models to optimize DTK's profits. We start by refining the notation for the models.

$$M(O, D, d) = M_A(O, D, d; t): \text{Airfare model}$$

$$M(Z) = M_H(Z; n, g, s, sm, h): \text{Hotel model}$$

where O: origin airport

D: destination airport

d: days between purchase of airline ticket and departure

t: departure date

Z: hotel zip code

n: # of nights in hotel

g: # of guests

s: # of stars of hotel

sm: smoking/non

h: average price on some date over last n years

Note that h = h(t), that is the average price on the date over the last n years depends on the date, t. With this notation, the optimization takes the following form

$$\max\left(\frac{gR}{g \times M_A(o,D,d\,;t) + M_H(z\,;n,g,s,sm,h)}\right) \quad \text{where } R \text{ is revenue per guest, } g.$$

$$\text{s.t.} \quad \begin{array}{l} o = o' \\ D = D' \\ t = t' \\ z = z' \\ n = n' \\ sm = sm' \\ h = h' \end{array} \qquad \left.\begin{array}{l} g_{min} \leq g \leq g_{max} \\[4pt] R_{min} \leq R \leq R_{max} \end{array}\right\} \begin{array}{l} \text{note that this is} \\ \text{a complicated} \\ \text{relationship which} \\ \text{should be determined} \\ \text{via a pricing model.} \end{array}$$

where g_{min}, g_{max}, R_{min}, and R_{max} are determined by choosing a range of values surrounding the optimum point as determined in the Supply and Demand Forecasting Model. Perhaps $g_{min} = 0.95 \times g_{opt}$, $g_{max} = 1.05 \times g_{opt}$ and similar for R_{min} and R_{max}. This wiggle room is allowed because a small shift from the forecasted g_{opt} and R_{opt} could potentially reduce the hotel cost more than it reduces the revenue (specifically, a small reduction in g may reduce the hotel cost more than it reduces revenue, g x R).

Summary

In total this system uses an A/B test, a meta-model of ARIMA models, a meta-model of regression models and a constrained optimization which leverages the models to maximize DTK's profits. The Supply and Demand Forecasting model as shown here assumes that the demand curve is linear and therefore only collects 2 data points. In practice, it may be better to map out multiple data points (test multiple costs) to more accurately represent the demand curve as it is likely non-linear.

These models would need historical data on airfare and hotel pricing. Deciding on a data source for airfare data is neither straightforward nor cheap as discussed here (https://www.quora.com/Are-there-any-free-APIs-for-flight-fare-search). Similarly for hotel data (https://www.quora.com/Which-is-the-best-source-for-statistical-data-on-hotel-occupancy). The ARIMA models underlying the Flight Purchase Optimizer, the regression model underlying the Hotel Purchase Optimizer and the Supply and Demand Forecasting module would all need to be re-trained periodically (maybe annually).