MIT SLOAN SCHOOL OF MANAGEMENT

15.093 OPTIMIZATION METHODS

MASTER OF BUSINESS ANALYTICS

Optimize to Survive

The Optimal Formula One 2023 Schedule

Students:
Guillaume Bonheure
Rodrigo Olivares-Lopez

Instructors:
Dimitris Bertsimas
Alexander Jacquillat

December 6, 2022

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Formula 1 has become one of the most popular sporting events worldwide, with a cumulative audience for the 2021 season of 1.55 billion people. By visiting a different city for each one its races, Formula One has become a truly global business with a tremendous economic impact. For example, the inaugural Formula One Miami Grand Prix benefitted the local economy to the tune of $350 million. Needless to say, what goes behind the planning of this traveling circus is extremely complex, to the point of having documentaries about it. Therefore, minimizing the cost and environmental impact of this logistical odyssey is not only lucrative but necessary. At the end of the day, a whopping 72% of Formula 1's 260K ton carbon footprint is created by its logistics and business travel. With all this in mind, optimizing the Race Schedule comes as one of the main challenges for the event organizers, and also a perfect opportunity for us to explore, apply, and expand on the knowledge covered in class, ultimately witnessing firsthand the edge of Optimization.

1.2 Problem Layout

The 2023 Formula One season will consist of a series of 24 races, known as Grands Prix, which take place worldwide on both purpose-built circuits and closed public roads. (For a list of the 24 events for 2023, please refer to Table 5 in Appendix A). As mentioned, the logistics of this global sporting event are complex; while a portion of each team goes back home whenever time permits (i.e. whenever there is more than a week between two races), another portion of team goes directly from circuit to circuit, bringing necessary equipment that is vital to each race's operations but does not need to return home between sub-tours (e.g. temporary living quarters, media equipment, etc.). Regardless, everyone is challenged by double and triple-headers (i.e. two or three consecutive racing weekends), where every member of the team needs to directly get to the next location in less than a week.

Our first approach to solve this problem turned out to be intractable: burdened by non-linearities, exponential constraints, and an overall abstract formulation. After regrouping, we decided to take this project as a “modeling journey”, starting from a simplified version of the problem and iteratively building on top of it. This document, in consequence, showcases the last iteration and latest milestone in our modeling journey. In summary, we decided to split the problem into two sub-problems that perfectly complement each other:

- **(1) Vehicle Routing Problem** to define the optimal sub-tours across the season; in other words, the double and triple-headers. Needless to say, this solution does not dictate the order in which these happen. Each time the ”vehicle” returns to the depot is an opportunity to choose which sub-tour to start next, as the order does not affect the objective value. This sub-problem corresponds to the ”Driving Team”: the portion of the team that goes back home as long as the next race is not the following weekend.
- **(2) Travelling Salesman Problem** to define the optimal order of the sub-tours and therefore define the actual schedule for the season. By imposing constraints representing the double and triple-headers (obtained from (1)). This sub-problem corresponds to the ”Logistics Team”, the portion of the team that does not go back home but rather carries out a season-long trip across the world and hence dictates the actual schedule for the year. As a reminder, the distance travelled by the Driving Team is the same for every solution of this Sub Problem, just in a different order.

For the sake of clarity, we present below the graphical solution to each one of these sub-problems and proceed to further explain how is that they complement each other.

As can be seen in Figure 1, the optimal solution to the Vehicle Routing sub-problem defines the optimal sub-tours: the single, double, and triple-headers. The single-headers, colored in grey, are those races that are not consecutive and allow the Driving Time to return home; therefore, the single bidirectional arc they have portrays this Home-Circuit-Home sub-tour they perform. On the other hand, the double and triple-headers, painted in different colors, represent the sub-tours that are made up of at least two consecutive racing weekends, where the Driving Team does not go back home in between. As mentioned previously, this solution does not impose an order on any of these sub-tours, as any possible combination leads to the same Objective Value.

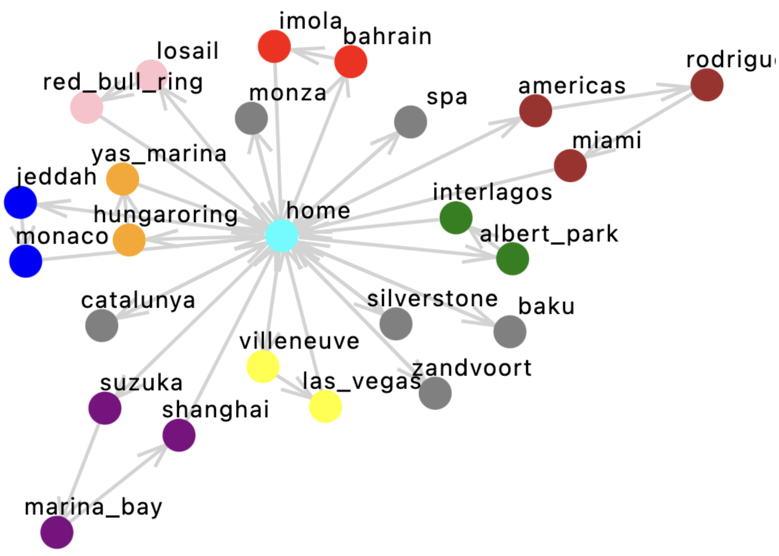

Figure 1: Optimal Solution to the VRP sub-problem

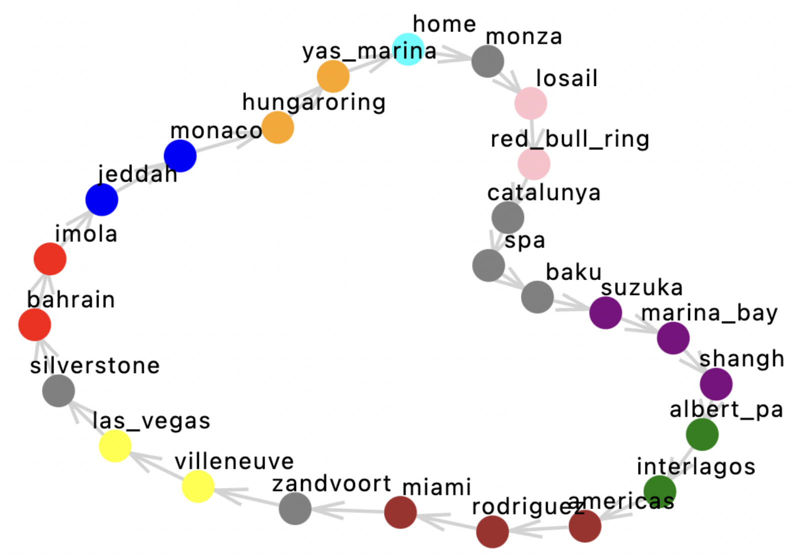

Figure 2: Optimal Solution to the TSP sub-problem

In parallel, Figure 2 shows the solution for the Travelling Salesman sub-problem. This solution portrays the optimal order for the sub-tours defined in the previous stage. More specifically, this is the tour that the Logistics Team takes around the world from venue to venue, regardless of whether the races are consecutive or not. As can be seen, the grey nodes still represent those races that do not happen consecutively, and all other painted nodes are colored with respect to the sub-tour they are part of. Ultimately, each change of color entails a trip back home for the Driving Team and the whole tour represents the Optimal Schedule for 2023.

2 Modeling

2.1 Data and Considerations

The data needed for this project consists of two main components: distances and costs, with the latter being both financial and environmental. To craft the problem as realistically as possible, we created a dataset with the coordinates of every circuit and every team's headquarters.

One of the main challenges we had at the beginning of the project was how to portray the fact that not all teams go back to the same depot (i.e. do not have the same home). After trying adding terms to the formulation and further increasing the computational burden, we took a step back and took a manual yet much simpler approach. We computed a distance matrix for each team and then took the average of all these matrices. By doing this, we are capturing in a single matrix the cost (i.e. distance) that the average team pays whenever they go back to their respective home/depot. This is a valid approach because all teams follow the same pattern of leaving and coming back to their headquarters at the same time. Hence, by multiplying the objective value by 10 (number of teams in Formula 1), we get the actual total distance travelled by all teams, taking into account the different location of each team's home.

While this distance matrix was a successful first approach that yielded significant results, we wanted to more realistically represent the impact of our solutions. Therefore, we decided to implement a Dual Objective formulation consisting of two measures of impact:

- **Emissions Matrix:** This Matrix is derived from the Distance Matrix but in a more comprehensive way, as we have modeled the fact that emissions are different depending on the mode of transportation. More specifically, by taking into account the distance between two venues and whether the team has only one week to get there (i.e. races are consecutive), each trip either offers the possibility to use trucks or requires the use of planes.
- **Attendance Matrix:** Optimizing for distance/emissions is a sensible approach; however, we realized that this entailed a potential trade off with profitability. If two consecutive races are held in locations not-too-far from each other, this could deter fans from potentially attending both races, yielding a loss in viewership and attendance. Therefore, we defined a matrix that characterizes a penalty on attendance based on how physically close two races are.

In more technical terms, the aforementioned matrices were computed in the following way, being slightly different for the VRP and TSP sub-problems:

- **Emissions Matrix VRP** We only discriminate on whether the distance to travel is less or greater than 5,000 km: the distance that can be realistically travelled by truck in one week. A truck emits around 62 grams per ton per km, while a plan emits around 500 grams per ton per km. Hence, the units of this matrix are [Emissions in Grams of C02 per Ton of Cargo].

$$e_{ij} = \begin{cases} d_{ij} \cdot 62 & if d_{ij} \leq 5,000km \\ d_{ij} \cdot 500 & if d_{ij} > 5,000km \end{cases} \tag{1}$$

- **Attendance Loss Matrix VRP** As discussed before, we impose a bigger penalty on those consecutive races that are held in venues too close to each other. Hence, the units of this attendance loss matrix are [Number of people]

$$a_{ij} = \begin{cases} 150,000 & if d_{ij} \leq 250km \\ 125,000 & if d_{ij} \leq 500km \\ 100,000 & if d_{ij} \leq 1,000km \\ 75,000 & if d_{ij} \leq 2,000km \\ 50,000 & if d_{ij} \leq 3,500km \\ 25,000 & if d_{ij} \leq 5000km \\ 0 & otherwise \end{cases} \tag{2}$$

- **Emissions Matrix TSP** Unlike the Emissions Matrix for the VRP, the one for TSP is constructed differently, as now there exists the possibility of having more than one week to get from location i to location j. Therefore, we establish a two-step restriction, assuming that, as long as two races are not consecutive, the Logistics Team has 12 days to get to the next venue. Whether the races are consecutive or not is recovered from the solution from the VRP.

  If races are consecutive, there only 5 days to move, so we enforce a 5,000 km limit to drive, otherwise teams must resort to the use of planes.

$$e_{ij} = \begin{cases} d_{ij} \cdot 62 & if d_{ij} \leq 5,000km \\ d_{ij} \cdot 500 & if d_{ij} > 5,000km \end{cases} \tag{3}$$

If races are not consecutive, there are up to 12 days to move, so we now impose a 8,000 km limit to drive, otherwise teams must resort to the use of planes.

$$e_{ij} = \begin{cases} d_{ij} \cdot 62 & if d_{ij} \leq 8,000km \\ d_{ij} \cdot 500 & if d_{ij} > 8,000km \end{cases} \tag{4}$$

- **Attendance Loss Matrix TSP** Finally, for this matrix we also recover the double and triple headers from solution of the VRP, and alleviate the attendance loss for races that are not consecutive: even if two circuits are physically close, if the races held in them are not-too-close to each other in the season, attendance does not suffer as much as if they were back-to-back, so their physical proximity should be penalized less.

$$a_{ij} = \begin{cases} a_{ij} & \text{if race i and race j are consecutive} \\ a_{ij}/2 & \text{if race i and race j are NOT consecutive} \end{cases} \tag{5}$$

2.2 Vehicle Routing Sub-Problem

The final formulation for the Vehicle Routing Sub-Problem took the following form:

$$\min \quad \sum_{i \in R} \sum_{j \in R} x_{ij} [\alpha \cdot e_{ij} + (1 - \alpha) \cdot a_{ij}] \tag{6a}$$

$$\text{subject to} \quad \sum_{i \in R} x_{ij} = 1 \quad \forall j \in R_1 \tag{6b}$$

$$\sum_{j \in R} x_{ij} = 1 \quad \forall i \in R_1 \tag{6c}$$

$$\sum_{i \in R_1} x_{i1} = K \tag{6d}$$

$$\sum_{j \in R_1} x_{1j} = K \tag{6e}$$

$$x_{ii} = 0 \quad \forall i \in R \tag{6f}$$

$$\sum_{i \in S} \sum_{j \in S} x_{ij} \leq |S| - 1 \quad \forall S \subseteq R_1, S \neq 0 \tag{6g}$$

$$x_{ij} \in \{0, 1\} \quad \forall i, j \in R \tag{6h}$$

$$doubleHeader_{ij} \leq x_{1i} \quad \forall i \in R_1, j \in [R_1 \setminus i] \tag{6i}$$

$$doubleHeader_{ij} \leq x_{ij} \quad \forall i \in R_1, j \in [R_1 \setminus i] \tag{6j}$$

$$doubleHeader_{ij} \leq x_{j1} \quad \forall i \in R_1, j \in [R_1 \setminus i] \tag{6k}$$

$$doubleHeader_{ij} \geq x_{1i} + x_{ij} + x_{j1} - 2 \quad \forall i \in R_1, j \in [R_1 \setminus i] \tag{6l}$$

$$\sum_{i \in R_1} \sum_{j \in R_1 \setminus i} doubleHeader_{ij} \leq DHL \tag{6m}$$

$$tripleHeader_{ijq} \leq x_{1i} \quad \forall i \in R_1, j \in [R_1 \setminus i], q \in [R_1 \setminus i, j] \tag{6n}$$

$$tripleHeader_{ijq} \leq x_{ij} \quad \forall i \in R_1, j \in [R_1 \setminus i], q \in [R_1 \setminus i, j] \tag{6o}$$

$$tripleHeader_{ijq} \leq x_{jq} \quad \forall i \in R_1, j \in [R_1 \setminus i], q \in [R_1 \setminus i, j] \tag{6p}$$

$$tripleHeader_{ijq} \leq x_{q1} \quad \forall i \in R_1, j \in [R_1 \setminus i], q \in [R_1 \setminus i, j] \tag{6q}$$

$$tripleHeader_{ijq} \geq x_{1i} + x_{ij} + x_{jq} + x_{q1} - 3 \quad \forall i \in R_1, j \in [R_1 \setminus i], q \in [R_1 \setminus i, j] \tag{6r}$$

$$\sum_{i \in R_1} \sum_{j \in R_1 \setminus i} \sum_{q \in [R_1 \setminus i, j]} tripleHeader_{ijq} \leq THL \tag{6s}$$

$$x_{1i} + x_{ij} + x_{jq} + x_{qr} \leq 3 \quad \forall i \in R_1, j \in [R_1 \setminus i], q \in [R_1 \setminus i, j], r \in [R_1 \setminus i, j, q] \tag{6t}$$

$$doubleHeader_{ij} \in \{0, 1\} \quad \forall i, j \in R_1 \tag{6u}$$

$$tripleHeader_{ijq} \in \{0, 1\} \quad \forall i, j, q \in R_1 \tag{6v}$$

Where:

- R is the set of the home {1} and the 24 circuits: {2,3,4,5, ... ,25}, so: {1,2,3, 4,5, ... ,25}
- R_1 is the set of just the circuits without the home: {2,3,4,5, ... ,25}
- x_{ij} is our decisions variable and determines whether the link between location i and j is active; if both locations are circuits (i.e. not home), this means they are consecutive races.
- e_{ij} is the ijth entry of the emissions matrix.
- a_{ij} is the ijth entry of the attendance loss matrix.
- $doubleHeader_{ij}$ determines if the arc (i, j) is part of a two-consecutive-races sub-tour.
- $tripleHeader_{ij}$ determines if the arc (i,j) is part of a three-consecutive-races sub-tour.
- S is the set that contains all subsets of R_1, in other words the powerset of R_1. This represents a set of all possible sequences.
- K is the number of sub-tours that we allow.
- Constraint 6a is the Dual Objective Function, depicting the trade-off between emissions and attendance
- Constraint 6b makes sure every node has only one incoming arc
- Constraint 6c makes sure every node has only one outgoing arc
- Constraint 6d and 6e make sure that there are exactly K sub-tours: the number of outgoing arcs is the same as the number of incoming arcs
- Constraint 6f makes sure there are no self-connecting sub-tours
- Constraint 6g ensures there are no sub-tours disconnected from the depot/home: the sum of all arcs within a potential sequence is at most the dimension of that sequence minus one. (e.g. for a potential sequence of three nodes, the active arcs cannot be more than two, otherwise a sub-loop could be formed between them)
- Variables x, doubleHeader, and tripleHeader are binary
- Constraints 6i to 6m limit the number of Double Headers in the solution, up to DHL.
- Constraints 6n to 6s limit the number of Triple Headers in the solution, up to THL.
- Constraint 6t ensures there are never more than three consecutive races.

2.3 Traveling Salesman Sub-Problem

The final formulation for the Traveling Salesman Sub-Problem took the following form:

$$\min \quad \sum_{i\in R}\sum_{j\in R} x_{ij}[\alpha \cdot e_{ij} + (1-\alpha)\cdot a_{ij}] \tag{7a}$$

$$\text{subject to} \quad \sum_{i\in R} x_{ij} = 1 \quad \forall j \in R_1 \tag{7b}$$

$$\sum_{j\in R} x_{ij} = 1 \quad \forall i \in R_1 \tag{7c}$$

$$\sum_{i\in R_1} x_{i1} = 1 \tag{7d}$$

$$\sum_{j\in R_1} x_{1j} = 1 \tag{7e}$$

$$x_{ii} = 0 \quad \forall i \in R \tag{7f}$$

$$\sum_{i\in S}\sum_{j\in S} x_{ij} \le |S| - 1 \quad \forall S \subseteq R_1, S \neq 0 \tag{7g}$$

$$x_{ij} \in \{0,1\} \quad \forall i,j \in R \tag{7h}$$

$$x_{ab} = 1 \quad \forall (a,b) \in \Gamma \tag{7i}$$

Where:

- Constraint 7i enforces the sequences coming from the optimal solution of the VRP. All pairs that make up the double and triple headers are enforced here. For example, the Triple Header Austin - Mexico - Miami makes the pairs (Austin, Mexico) and (Mexico, Miami) to be part of Γ.
- Constraints 7d and 7e make sure there is only one sub-tour, i.e., we have a Travelling Salesman formulation that departs from the depot, visits each node only once, and returns back home.

3 Baseline

In order to fully appreciate the edge of Optimization, we needed a baseline to compare against. For this, we recovered the official calendar for 2023 and calculated the objective value that this ”solution” brings about.

With the same reasoning, we computed two objective values, one for the corresponding VRP (portion of the team that goes back home) and another one for the TSP (the rest of the team that travels from circuit to circuit). The corresponding emissions and attedance loss matrices were also generated and taken into account. A visual representation of both sub-problems for the baseline solution (i.e. Official 2023 Schedule), can be found in Figures 5 and 6 in Appendix A.

4 Results

Given the number of constraints that make up the Vehicle Routing Problem ($\approx$17 Million, mainly due to the subtour elimination constraints), our personal computers were not able to handle constructing the model for more than 20 races. Hence, we decided to run this model using MIT's cloud computing services. A node with 24 cores and 512 GB of RAM was able to solve each iteration of the VRP in *practable* time, but only after implementing the improvements detailed below. The formulations presented in this document were implemented, letting α iterate between 0.0 and 1.0 in 0.05 increments. Also, DHL, the limit on Double Headers, was set to 6, and THL, the limit on Triple Headers, was set to 2, all in accordance with the Status-Quo in the sport.

We realized that there was a way to decrease the number of constraints needed for the Double and Triple Headers. Given the fact that we knew the number of sub-tours that the optimal solution would take (basically a binding constraint on the double and triple headers), a parallel approach was to impose a restriction on the number of single-headers. Hence, constraints 6i to 6s from the VRP could be replaced with the constraints below, leading to the same optimal solution in a significantly more efficient way as the amount of variables decreased from more than 14K to just 650. The amount of Single Headers is fixed to 6, given by the fact that the optimal solution creates as many Double and Triple headers as it can. Therefore, out of the 24 races, the binding constraints for the Double and Triple Headers will occupy 18 races (6 Double headers and 2 Triple Headers), leaving only 6 individual races. More generally: K = DHL + THL + SH, where SH is the number of Single Headers and is defined simply by subtracting the races that are part of double and triple headers from the total amount of 24 races: SH = 24 - 3 * THL - 2 * DHL = 6. And hence K = 14.

$$singleHeader_i \leq x_{1i} \quad \forall i \in R_1 \tag{8a}$$
$$singleHeader_i \leq x_{i1} \quad \forall i \in R_1 \tag{8b}$$
$$singleHeader_i \geq x_{1i} + x_{i1} - 1 \quad \forall i \in R_1 \tag{8c}$$
$$\sum_{i \in R_1} singleHeader_i = SH = 6 \tag{8d}$$

This small change brought the computation time for each iteration to about three hours, enabling us to experiment more easily with the model. Lastly, both as a heuristic to decrease computational time and as a realistic improvement, both sub-problems were ran with Abu Dhabi fixed as the closing event of the season. Enforcing this constraint removes a degree of freedom and further alleviates the computational burden; furthermore, the Abu Dhabi Grand Prix has signed a 10-year contract with the exclusive rights to the ending race of the season. The index corresponding to Abu Dhabi is 25.

$$x_{25,1} = 1 \tag{9a}$$
$$x_{25,j} = 0 \quad \forall j \in R_1 \tag{9b}$$

4.1 Vehicle Routing Problem

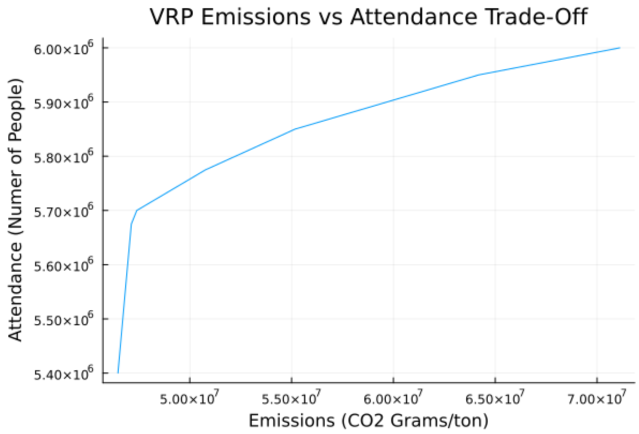

Figure 3: Pareto Curve for VRP Problem

Priority	alpha	Distance [km]	Emissions [grams/ton]	Attendance [People]
Reduce Emissions	0	139,810	46.47 M	5.4 M
Increase Attendance	1.0	179,262	71.11 M	6.0 M
Sweet Spot	0.1	154,725	47.39 M	5.7 M
Baseline	NA	189,508	74.95 M	5.4 M

Table 1: VRP - Dual Objective Results

After being able to analyze the trade-off between Emissions and Attendance, we decided to select the Solution with $\alpha = 0.1$ as the Optimal ”Sweet-Spot” Solution, since at this point the trade-off for increasing attendance any further comes with massive increases in emissions, and vice-versa decreasing emissions a bit would lead to a massive attendance decrease. It must be noted that, in order to make the Pareto Curve more interpretable from a business perspective, we plotted Attendance instead of Attendance Loss. This solution improves on the baseline in every aspect: Distance Travelled, Emissions Generated, and Attendance. The graphical representation of this solution is shown in Figure 1 at the beginning of the document. More specifically, we see the following improvement for the main team that goes back home:

Distance Reduction	Emissions Reduction	Attendance Improvement
18.35 %	36.77%	5.55%

Table 2: VRP - Improvement over Baseline

4.2 Travelling Salesman Problem

By the same token, the ”sweet spot” optimal solution from the VRP that minimizes attendance loss without significantly increasing emissions was selected in order to compute the TSP. The computer setup in the cluster managed to solve each iteration in about 30 minutes also varying α between 0 and 1 in 0.05 increments.

Priority	alpha	Distance [km]	Emissions [grams/ton]	Attendance [People]
Reduce Emissions	0	96,507	10.76 M	5.35 M
Increase Attendance	1.0	133,713	25.64 M	5.65 M
Sweet Spot	0.35	110,600	11.63 M	5.55 M
Baseline	NA	131,865	37.68 M	5.25 M

Table 3: TSP - Dual Objective Results

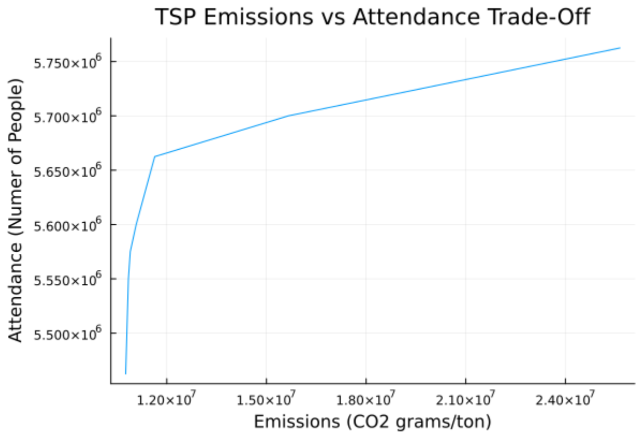

Figure 4: Pareto Curve for TSP Problem

In this case, we also were able to easily identify a "Sweet-Spot" optimal solution for $\alpha = 0.35$, where once again a small increase in attendance would lead to a significant increase in emissions and analogous for decreasing emissions. Intuitively, this could be interpreted as being able to maintain a strong viewership and attendance by making slightly less-compact sub-tours, yielding more time in between nearby circuits to allow for fans' excitement to build up. The visual representation for this solution can be seen in Figure 2 at the beginning of the document. This solution is the optimal 2023 Formula One Schedule. Just like in the VRP sub-problem, this solution improves over the baseline in every aspect.

Distance Reduction	Emissions Reduction	Attendance Improvement
16.12 %	69.13%	5.7%

Table 4: TSP - Improvement over Baseline

4.3 Conclusion

Defining a formulation to realistically represent the logistical problems of Formula One proved to be a challenging task; nevertheless, going through several modeling iterations allowed us to decompose the formulation in insightful ways. Namely, by separating the "Logistics" and "Driving" teams into separate yet complementary problems, allowed us to witness the edge of optimization from different angles and in different stages of the decision making process. One intractable problem could be successfully decomposed into two tractable sub-problems that yielded more flexibility and also provided more insights for each component of the solution.

More importantly, we can see that any of our solutions increases attendance and decreases emissions, pointing in the direction of how much improvement can be done over the baseline. Even in the worst case scenario in which our emissions calculations are not too realistic, the distance reduction is an objective measure that would scale accordingly to the true impact. Interestingly enough, we can see how one of our solutions for the TSP (the one that prioritizes attendance) has more distance travelled yet lower emissions, this happens because we allow for more time to do the long travels, which means we do not have to resort to the use of planes as often. This is a very insightful observation that only optimization would had been able to shed light on, as it seems counter-intuitive on the surface.

Finally, we could see and experience how is it that some domain expertise can significantly reduce the complexity of a formulation without sacrificing analytical rigor. If knowledge regarding the problem allows for realistic simplifications and assumptions, these can become powerful modeling tools, just like we realized with the substitution of the Double and Triple Headers with a simpler set of constraints for Single Headers. Without a doubt, it was a fruitful project that paved the way for a deeper understanding of Integer Programming and Combinatorial Optimization but, most importantly, the modeling journey behind tackling real-world problems through Optimization.

5 Appendix

5.1 Appendix A

Date	Grand Prix	Venue
February 23-25	Pre-season testing	Sakhir
March 5	Bahrain	Sakhir
March 19	Saudi Arabia	Jeddah
April 2	Australia	Melbourne
April 16	China	Shanghai
April 30	Azerbaijan	Baku
May 7	Miami	Miami
May 21	Emilia Romagna	Imola
May 28	Monaco	Monaco
June 4	Spain	Barcelona
June 18	Canada	Montreal
July 2	Austria	Spielberg
July 9	United Kingdom	Silverstone
July 23	Hungary	Budapest
July 30	Belgium	Spa
August 27	Netherlands	Zandvoort
September 3	Italy	Monza
September 17	Singapore	Singapore
September 24	Japan	Suzuka
October 8	Qatar	Lusail
October 22	USA	Austin
October 29	Mexico	Mexico City
November 5	Brazil	Sao Paulo
November 18	Las Vegas	Las Vegas
November 26	Abu Dhabi	Yas Marina

Table 5: 2023 Official Schedule

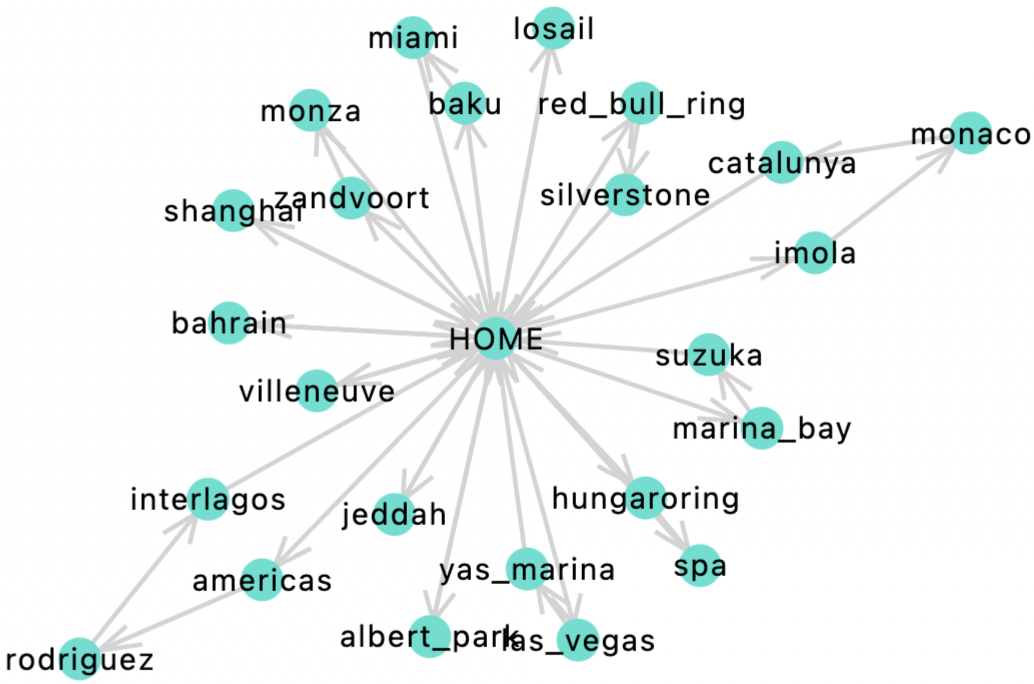

Figure 5: Baseline Vehicle Routing Problem - Official Schedule

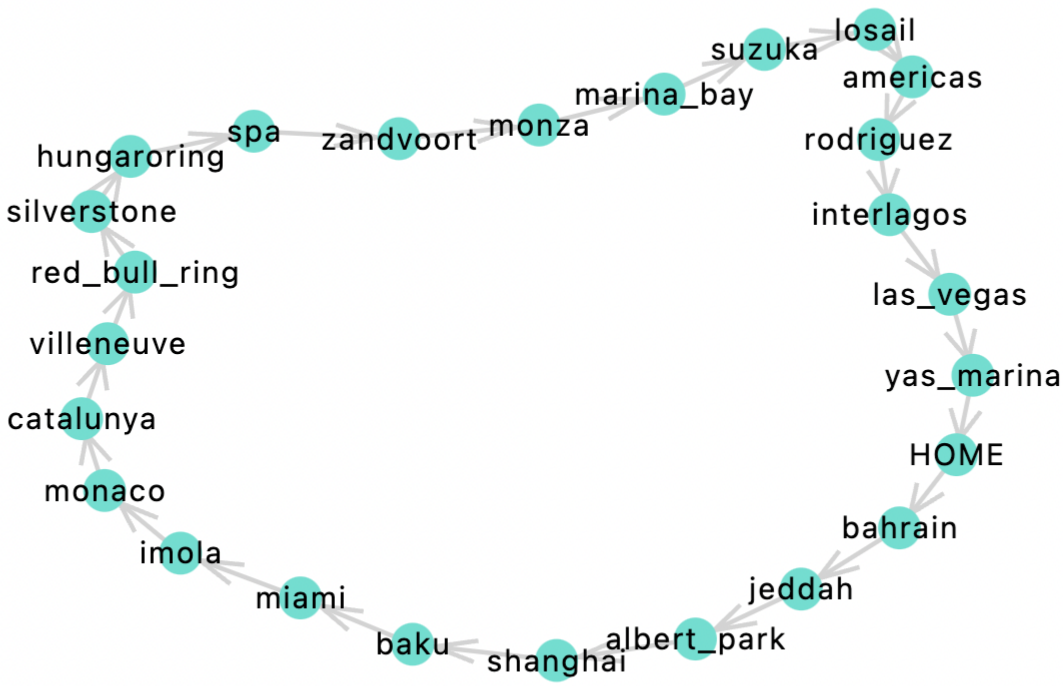

Figure 6: Baseline Travelling Salesman Problem - Official Schedule