ChatGPT

Economic Impact Estimation Model for Endurance Running Events

Introduction

Endurance running events such as marathons and half-marathons can draw thousands of participants and spectators, generating significant spending in host communities. To capture this economic contribution, we propose a **flexible, Python-compatible economic impact model** that quantifies both direct and secondary (indirect and induced) impacts of an event. The model prioritizes accuracy by using actual post-event survey data when available, and applies credible proxy values when data are missing. It accounts for key spending categories – **accommodation, daily expenditures (food, entertainment, etc.), transportation, and event organizational outlays** – focusing on **new money brought in by non-local visitors** while excluding most routine local spending [1]. The goal is to estimate total direct spending in the host economy, the ripple effects (indirect and induced impacts), and associated tax revenues [2] in a consistent, automated manner.

This report outlines the model's data requirements and schema, the logic for switching between actual and proxy data, calculation methods for direct and secondary impacts, and how the model handles uncertainty. We also describe how the model's output can be generated as a structured report (mirroring the layout of the 2024 Shamrock Marathon Economic Impact Report) and as machine-readable metrics, facilitating integration into a future UI/UX pipeline.

Data Inputs and Schema

To ensure flexibility, the model accepts **structured input data** either from user-provided files (e.g. spreadsheets or survey CSV) or API payloads. **Table 1** defines the recommended schema for the key input fields, along with descriptions and default proxy assumptions if actual data is unavailable. This schema is designed to capture event-specific parameters such as size, location context, attendance breakdown, and spending profiles.

Table 1. Key Input Fields and Proxy Assumptions

Input Field	Description	Data Source / Proxy (if missing)
Event Name / Date	Identifier and date of the event (for report context).	*Required.* Used in report headers and context.
Location & Type	Host city/region and context (urban, rural, tourist hub).	*Required.* Used to select appropriate spending proxies and multipliers. A "tourist destination" may assume longer stays and higher spend per visitor than a local rural event.

Input Field	Description	Data Source / Proxy (if missing)
Total Participants	Total number of race finishers/participants.	*Required.* Basis for all calculations (often known from registration data).
Non-local Participant %	Percentage of participants from outside the host region.	If not provided, estimate from comparable events or registration zip code data. E.g. use past surveys or studies (53% nonlocal for similar marathons [3]). If unavailable, default to ~50% for large city marathons or ~20% for small local races, adjustable by location.
Total Spectators	Number of event spectators/visitors (excluding participants).	If actual count is unavailable, model will estimate based on travel party size or a spectators-per-runner ratio. E.g. assume each non-local runner brings ~1–2 additional people on average [4], and local runners might bring some local spectators (excluded from economic impact unless they are non-local visitors).
Average Travel Party Size	Average number of people in each travel group (including the participant).	Use survey data if available (e.g. 2.7 people per party including the runner [4]). If unknown, use benchmarks: large destination marathons might average ~2–3, smaller regional races ~1.5–2 [5]. Split by context: travelers from far away tend to have larger parties than day-trippers [5].
Overnight Stay %	Proportion of non-local visitors who stay overnight (vs. day trips).	If not given, infer from travel distance or event duration. For example, classify attendees by distance: those traveling >~1.5 hours are assumed to stay overnight [6]. For a multi-day event or a tourist city, assume higher overnight percentage.
Average Nights (Length of Stay)	If overnight, the average number of nights stayed.	From surveys if possible (e.g. ~2.5 nights [7] or 1.7 nights for those >1.5h away [8]). If unknown, assume 2 nights for a standard marathon weekend (adjust upward in a tourist destination or for multi-race weekends, downward if most live nearby).
Average Daily Spending per Person	Average amount an individual visitor spends per day on *food, entertainment, local transport, and shopping* (excluding lodging).	Use actual survey results if available. If not, use industry benchmarks: studies show marathon visitors often spend around \$150–\$200 per person per day on food, transport, and entertainment [4] [9]. This can be refined by context (higher in big cities or tourist areas).

Input Field	Description	Data Source / Proxy (if missing)
Average Lodging Rate	Average accommodation cost per night (per room).	From survey or local hotel data (e.g. reported ADR – average daily rate – in host city). If unknown, use city's average hotel rate or a typical rate (e.g. \$120–\$200/night depending on locale [10]). Include vacation rentals if significant.
Rental Car Utilization	Percentage of visitor groups that rent cars and average rental cost.	Use survey data (e.g. what % flew in and rented a car, and cost). If not available, estimate a small fraction for urban events with good transit, higher for destinations where most arrive by air. Average cost can be assumed (e.g. \$50–\$100 per day for those who rent).
Local Organizational Spending	Amount the event organizers spend locally (suppliers, staff, services) and any local revenue retained.	If the event budget is known, include the portion spent in the host economy (e.g. local vendors, staff wages). Surveys or financial reports may provide this. If unknown, estimate using similar events' budgets (for instance, large marathons often spend \$3–\$5 million locally [11]). In absence of data, a proxy of \$X per participant (e.g. \$50–\$100) can be used to approximate organizer spending.
Tax Rates	Relevant local tax rates (sales tax, hotel occupancy tax, etc.).	Provide if available (e.g. sales tax 8%, hotel tax 10%). Otherwise, use typical local rates to compute tax revenues from spending.

Data Formats: These inputs can be supplied in a **structured spreadsheet or JSON**. For example, a CSV or Excel template might have fields for the above parameters. The model could also accept a JSON object via API, for example:

```
{
  "event_name": "City Marathon 2025",
  "location_type": "Urban",
  "total_participants": 5000,
  "percent_nonlocal": 60,
  "average_party_size": 2.5,
  "overnight_stay_percent": 70,
  "average_nights": 2.0,
  "avg_daily_spend_per_person": 160.00,
  "avg_lodging_rate": 150.00,
  "rental_car_percent": 20,
  "avg_rental_cost_per_day": 60.00,
  "organizer_local_spend": 500000,
  "sales_tax_rate": 0.08,
```

```
  "hotel_tax_rate": 0.10
}
```

When detailed **survey data** is available (e.g. individual responses on travel party, nights, spending), the model can ingest it (as a CSV of responses) and internally compute the above aggregate inputs. For instance, the Shamrock Marathon survey collected each participant's stay details and expenses [12] [13]; our model would summarize such data to derive the average party size, average nights, spending per person, etc. If a survey is provided, it takes precedence over default values to ensure accuracy.

Logic for Actual Data vs. Proxy Data

The model is designed to **dynamically switch** between using actual event-specific data and proxy assumptions depending on availability:

- **Survey-Driven Mode (Actual Data):** If post-event survey results or detailed attendee data are provided, the model parses these to obtain precise inputs. For example, it will calculate the exact average travel party size, nights stayed, and spending patterns from the responses and extrapolate to the full participant population [13]. These values replace any default assumptions. The model also uses the survey to determine splits like what percentage of runners were local vs. visitors, how many spectators traveled, etc. (For instance, if 53% of participants are non-local in a similar marathon [3], that figure would be used directly for that event if applicable.) All calculations then proceed from these empirically grounded inputs.

- **Proxy Mode (No/Partial Survey):** When survey data is unavailable, incomplete, or only some inputs are known, the model falls back on a library of **benchmark data and assumptions**. This library is built from prior economic impact studies and industry research, ensuring the proxies are reasonable. Key fallback logic includes:

- *Non-local percentage:* If not known, use historical data from comparable events or studies (e.g., smaller regional races might assume ~50% outsiders, whereas destination races often see >80% outsiders [4]). The model can adjust this by event location type (higher for popular destinations, lower for community races) [14].
- *Spectators:* In absence of a measured spectator count, estimate via **travel party size**. The model assumes each non-local participant brings some companions (friends/family), contributing to spectator count [4]. It also accounts for **non-local spectators visiting locals** – e.g. if local runners have family traveling in for the race, those visitors count as new economic contributors (captured via travel party inputs or a separate estimate if provided). Pure local spectators (from the host city) are *not counted* as they don't inject new funds, consistent with standard practice [15] [16].
- *Overnight vs. day-trip split:* Use distance or event duration as a proxy. For example, the model can take participants' origin data (if available from zip codes) to classify who likely needed lodging: those traveling beyond a threshold (like >90 minutes) are assumed to stay overnight [6]. In one marathon study, anyone over ~1.5 hours away was counted as an overnight visitor [17]. If origin data isn't provided, the model can default to an assumed percentage of overnight visitors (tailored by event type and location). Multi-day events or those in resort areas will have a higher assumed overnight rate than one-day events in non-tourist towns.

- *Travel party size:* If not measured, use averages from similar events or tourism data. Research suggests travel parties for marathons typically range from about 2 to 3 people including the runner [4] [5]. The model may use a default like 2.0–2.5 people per party for most events, and can increase that for family-oriented destination races (or decrease for local-focused races). Additionally, it can differentiate day-trip parties (often just the participant or one companion) versus overnight parties (which tend to be larger) [5].
- *Spending per person:* In absence of event-specific spending data, the model draws on **benchmark spending profiles**. It will use prior studies to estimate daily spending in categories such as lodging, food, transport, and entertainment [9]. For example, an **adult marathon overnight visitor** might spend roughly \$150–\$200 per day (per person) on non-lodging items [18], whereas a day-trip attendee spends less (since no lodging and possibly fewer meals) [18]. The model will adjust these numbers for the local price level (a city with higher costs or tourist pricing might be at the upper end of the range). If only total trip spending estimates exist, the model will convert to per-day values given the length of stay.
- *Lodging rate and nights:* Use average hotel rates in the region if not provided. Many economic impact reports include the average daily rate (ADR) for hotels during the event [10] – if the user doesn't supply one, the model can fetch or assume a typical rate based on the city's data (potentially via integration with tourism stats or user input). Nights are set as described (usually 1–3 nights).
- *Organizer spending:* If not given, the model can **proxy the event's local expenditures** by using either industry heuristics (like cost per runner, or known budgets from similar races) or leaving it as zero for a conservative estimate. Since organizer spending can be a significant direct impact (e.g., Grandma's Marathon's organizers spent \$3.5M locally [11]), it's ideal to include at least an estimate. The model might default to a certain percentage of total direct spending or ask for an estimate if not provided.

The logic is implemented in a **decision tree** within the Python code. For each input, the code checks: if a value is provided or derivable from provided data, use it; if not, apply the corresponding proxy rule. The model will document in the output which inputs were actual vs. assumed, for transparency. This flexible design ensures the model produces a result even with minimal data, while automatically favoring real data to improve accuracy [19] [14].

Direct Economic Impact Calculation

Once the input parameters are set (from either actual or proxy data), the model computes the **direct economic impact** – the total spending in the local area directly attributable to the event. In line with standard economic impact analysis, **only spending by visitors (non-local participants and their parties)** is counted, since local spending is not new to the economy [1] [15]. The direct impact is calculated by summing spending across several categories:

- **Accommodation (Lodging):** This includes hotel nights and other lodging (vacation rentals, campgrounds, etc.) used by visitors. The model calculates total lodging revenue as:

  $$Lodging\ Spending = Overnight\ Visitors \times Average\ Nights\ Stayed \times Average\ Nightly\ Rate.$$

  Here, *Overnight Visitors* is derived from the number of non-local participants and their accompanying guests who required lodging (based on the overnight percentage or travel distance logic). For example, if 10,000 total visitors stayed an average of 2 nights at \$150/night, lodging direct spending = \$3,000,000. This output can be further broken down in the report (e.g. showing total room-nights and the ADR) [20] [10]. If some visitors stay with friends/family, those are not counted as they don't generate local lodging revenue (survey data can reveal what portion stayed in paid accommodations).

- **Daily Expenditures:** This category captures all **non-lodging daily spending** by visitors, such as food and beverage (restaurants, bars), local transportation (taxis, rideshare, gas, parking), entertainment (attractions, nightlife, sightseeing), and shopping (retail purchases, souvenirs). The model computes this by first finding the total number of **visitor-days** (each visitor multiplied by the number of days they are in town) [16]. Then:

$$Daily\ Expenditures = Total\ Visitor\text{-}Days \times Avg.\ Daily\ Spend\ per\ Person.$$

  If separate average spends are known for different categories, the model can sum them up; otherwise a single aggregated daily spend is used. For example, using the earlier party example: 10,000 visitors staying 2 days = 20,000 visitor-days; if each spends \$70/day on food, \$20 on transport, and \$30 on entertainment/shopping (total \$120/day), that yields \$2.4 million in daily expenditures. The model distinguishes overnight visitors and day-trippers if needed, as day-trippers might have lower daily spend (no dinner or less leisure spending) [21] [22]. All such calculations can be tuned with survey data – for instance, Blue Ridge Marathon data showed overnight parties spent more on meals and gas than day-trip parties [21]. The final output will report the total **Daily Expenditures Contribution** from out-of-town visitors [23].

- **Transportation (Air & Ground):** This can be split into two parts: **rental cars/ground transport** and **air travel**. Typically, **airfare** spending is *excluded* from local impact since most of that revenue goes to airlines (unless the host city captures a portion via airline fees). However, **local transit and car rentals** are included. The model calculates rental car spending if applicable:

$$Rental\ Car\ Spending = Visitor\ Parties\ Renting\ Cars \times Avg.\ Rental\ Cost\ per\ Trip.$$

  This might use daily rental cost × days rented. Ground transport like Uber, taxis, gas, and parking are usually counted within daily expenditures (transport sub-category). If survey data gives a combined figure for "ground transportation" per person [24] [25], the model will include that in daily spend. In the output, we may present a separate **Rental Car Contribution** line if it's significant or calculated distinctly [23]. For example, the Shamrock report listed about \$76k from rental car spending [26], reflecting relatively small share, whereas food and lodging were much larger.

- **Event Organizational Spending:** This captures the **expenditures by the race organizers in the local area**. Any spending on local vendors, contractors, facilities, staff wages, and services for the event contributes directly to the economy. Additionally, any event profits that remain local (or portions of registration fees that go to local charities, for instance) can be counted. The model will take the **Organizer Local Spend** input and include it fully as direct impact. If we have details, we can break it down (e.g. spending on operations vs. local charities vs. local staff). In the absence of precise data, this is one area of uncertainty – but it can be significant. For example, organizer spending contributed roughly \$3 million in one marathon's impact [23]. Our model is prepared to incorporate this either via actual budgets or estimates.

Summing the above components yields the **Total Direct Spending** attributable to the event. In formula form:

$$Total\ Direct\ Impact = Lodging + Daily\ Expenditures + Transportation + Organizer\ Spend.$$

It's important to note that **local participants' spending is generally excluded** (aside from possibly their race entry fees if those go to the organizer and are spent locally, or any spending at event-specific vendors). This prevents counting money that would likely have been spent locally anyway [15]. The model follows this principle, aligning with industry standards for economic impact analysis. Any included local spend (like a local runner buying gear at an expo from an out-of-town vendor) is handled cautiously and would be a minor component if at all [27].

Indirect and Induced Impact Calculation

Direct spending by visitors is only one part of the economic effect. The model also estimates **indirect and induced impacts** – the secondary "ripple" effects as that initial spending circulates through the local economy. We support two approaches to calculate these effects:

- **Multiplier-based estimation:** The simplest method is to apply an **economic multiplier** to the direct spending. The model can use a single "total output multiplier" or separate multipliers for indirect and induced effects. These multipliers are typically obtained from regional input-output models (like IMPLAN or RIMS-II) or provided by local economic development agencies. For example, if a suitable tourism multiplier is 1.5, and direct spending is \$10 million, then total economic impact = \$10m × 1.5 = \$15 million, implying \$5 million of indirect+induced effects. If more granular data is available, different multipliers can be applied to different categories (e.g. lodging might have a different multiplier than retail). By default, the model can store typical multipliers by industry for the region or use an aggregate one if detailed data is lacking. These multipliers capture how, for instance, hotels purchase supplies and hire staff (indirect effects) and how workers spend their incomes locally (induced effects) [28] [29].

- **Integration with Input-Output Models:** For advanced usage, the model can integrate with or export data to specialized economic impact software (such as IMPLAN or an open-source IO model). In the Shamrock Marathon analysis, for example, the researchers fed spending data into IMPLAN Pro 3.0 to compute indirect and induced impact [30]. Our Python model can be designed to output a formatted spreadsheet of spending by category (lodging, food, retail, etc.), which an analyst could input into IMPLAN. Alternatively, if simplified multipliers are known (e.g. IMPLAN's reported total indirect+induced = \$X for a given direct input), the model can incorporate those. The key is that **indirect impact** represents business-to-business supply chain effects (e.g. restaurants restocking food, as defined in the Shamrock report [31]) and **induced impact** represents the household spending of incomes supported by the event [32].

By default, our model will apply a **combined multiplier** to produce an estimate of **Total Indirect & Induced Impact**, which can then be split into indirect and induced if needed (say, 40% indirect / 60% induced based on typical breakdowns, or using known ratios from studies). For example, if direct spending

= \$10M and we assume indirect+induced adds ~50%, we'd estimate \$5M additional, split perhaps into \$2M indirect + \$3M induced. These ratios can be refined if the user provides specific multipliers. We will clearly document the multiplier used or source. In the output, this is reported as **Indirect & Induced Impact** [23], and when added to direct impact gives the **Total Economic Impact**. This approach follows the accepted economic impact formula:

$$Total\ Impact = Number\ of\ Visitors \times Avg.\ Spending\ per\ Visitor \times Multiplier\ 【11†L85-L93】.$$

Using appropriate multipliers ensures we measure the **"ripple effect"** of runner and visitor spending through the economy [30]. The model can store default multipliers by region type: for instance, a smaller local economy might have a lower multiplier (more leakage) whereas a large city retains more spending (higher multiplier). If available, a *capture rate* (the portion of spending that remains in the local economy after leakage) can also be applied for more accuracy [34].

Tax Revenue Estimation

In addition to spending impacts, many stakeholders care about **fiscal impacts** – how much tax revenue the event generates for local government. Our model will calculate **tax revenues** based on the direct spending components and provided tax rates. Specifically, it will estimate:

- **Sales Tax** from taxable sales (e.g. meals, shopping, local transport). For example, if out-of-town visitors spent \$5 million on taxable goods/services and the sales tax rate is 8%, then \$400,000 in sales tax is generated.
- **Hotel Occupancy Tax** (and other lodging-related fees) from accommodation spending. For instance, a 10% hotel tax on \$3 million in lodging spend would yield \$300,000 for the city [35].
- **Rental Car Tax/Fees** if applicable (some locales have special taxes on rental cars) [36].
- Any other event-specific taxes or fees – for example, the Shamrock report even noted a small amount of **airport/landing fees** due to event-related travel [37], which our model can include if relevant inputs are given.

The model calculates these by applying the **user-specified tax rates** to the corresponding spending totals. All taxes are then summed to report **Total Taxes & Fees Generated** [35]. If local tax rates are not provided, the model can use typical rates for the jurisdiction (though for accuracy, user input is encouraged). These figures will be included in the structured report, often in a section detailing tax benefits. For example, the output might list Sales Tax, Lodging Tax, etc., and then a total tax revenue number. This helps cities justify the event by showing a return in tax dollars. (Note: These tax revenues are usually a subset of the direct spending, not additional economic impact, so they are reported separately rather than added on top of the total impact).

Output: Structured Report and Metrics

The model's output is twofold:

1. **Comprehensive Report (Automated):** A narrative report is generated, closely mirroring the structure and tone of professional economic impact studies (like the provided 2024 Shamrock EI Report). It will include clearly labeled sections such as: **Introduction/Background** (event description and objectives of analysis), **Methodology** (data sources, survey info or proxy use, and how

calculations were done), **Economic Impact Findings** (detailed results), and possibly appendices for technical details or survey questions. The Findings section will present the key metrics – total direct impact, breakdowns by category, indirect/induced impact, and tax revenues – in a readable format, often with bullet points or tables for clarity. For instance, it may have a summary like: "**Total Economic Impact:** \$X million, composed of \$Y million direct spending and \$Z million in indirect/induced effects." Then it will detail **Accommodation Contribution, Daily Expenditures Contribution, Rental Car Contribution, Organizational Contribution,** etc., each with the dollar amounts and an explanation [23]. Where appropriate, simple tables or bullet lists will be used to make the data easy to scan (e.g. a table of "Hotel Room Nights and Average Rate" in a hotel section, or a bullet list of key findings in an executive summary). The language will be non-technical and aimed at stakeholders, while an appendix can contain methodological notes (like assumptions made, survey response rates, etc.). Because this report is generated from templates, it ensures consistency across events while allowing customization (like inserting the event name, dates, and city specifics in the narrative).

2. **Numeric Output (Data Tables/JSON):** Alongside the human-readable report, the model produces structured data of the results for use in other applications. This could be an Excel sheet or JSON object containing all relevant metrics. **Table 2** outlines the key output metrics that the model will return. These metrics can feed into dashboards, comparison analyses, or be stored in a database. If the model is accessed via API, the JSON response would contain these fields. If through a spreadsheet, a tab might list these summary figures.

Table 2. Output Metrics Definitions

Metric	Description
Total Direct Spending	Total new spending in the local economy by event visitors and organizers (sum of lodging, daily expenditures, transport, etc.). This is the primary direct economic impact of the event.
Accommodation Spending	Direct spending on hotels and other lodging by out-of-town participants and their parties. Often presented separately as it's a significant category [23]. Includes room nights × ADR (average daily rate).
Daily Expenditures	Total spent by visitors on daily needs (food, beverages, entertainment, local transport, shopping) during their stay. This aggregates all per-person daily spending across all visitor-days.
Transportation Spending	Direct spending on local transportation services, which may include rental car revenue (or listed separately), fuel, public transit, taxis/rideshares related to the event visitors. (Air travel costs are generally excluded or noted separately if considered.)
Organization Spending	Local expenditures by the event organizers and any locally retained revenues (sponsorship money spent in-region, supplies bought locally, staff salaries, charity donations in the area, etc.).

Metric	Description
Total Indirect & Induced	The secondary economic impact generated through supply chain purchases and employees' re-spending of income [24] [29]. This is usually calculated via multipliers. It may be broken into two components (indirect and induced) in the report or supporting data.
Total Economic Impact	Sum of direct, indirect, and induced impacts – the overall contribution of the event to the local economy. This is the headline figure often reported (e.g. "Event X generated \$Y million in economic activity"). It encompasses all rounds of spending [33].
Tax Revenues	Total local taxes and fees generated by the event-related spending. This can be subdivided into sales tax, lodging tax, etc., depending on what is relevant and available [35]. Only the taxes accruing to the local region are counted (state or federal taxes can be noted but are usually outside the local impact scope).
Ancillary Metrics	Any additional outputs, such as **Total Visitors** (number of non-local attendees including spectators), **Total Visitor-Days**, **Hotel Room Nights**, **Average Spending per Visitor** (for benchmarking), and even jobs supported (if an employment impact conversion is applied from the indirect/induced analysis). These provide context and can be included in the report narrative or appendix.

The model ensures that all these metrics are internally consistent (e.g., the lodging spending corresponds to the reported room nights and ADR, etc.). The **structured data output** is crucial for automation – it allows a front-end interface to easily grab the numbers and populate visuals or compare multiple events. For instance, an interactive dashboard could use the JSON to display a chart of direct vs. indirect impact, or a summary across years.

In the generated **report document**, these metrics are woven into sentences and tables for readability. For example, a paragraph might read: *"Total direct spending by visitors for the marathon weekend is estimated at \$13.1 million, including \$3.27M on accommodation, \$6.68M on food, entertainment and other daily expenses, \$76K on rental cars, and \$3.07M in local event organization spending [38]. This direct injection of spending led to approximately \$9.06M in additional indirect and induced economic activity, for a total economic impact of \$22.16M in the region [23] [38]."* Each figure in the text is sourced from the model's calculations. The report might also include a brief **Executive Summary** with key figures (especially useful for stakeholders who want a quick overview), as well as charts or bullet lists highlighting important points (for example, a bullet list of *key findings* like "X% of participants were from out of town", "Average spending per visitor was \$Y", etc., if such insights are available from the data).

Importantly, the report will **document assumptions** when proxies are used. For instance, if no survey was available, a methodology note would state: "Average daily spending of \$160 per person was assumed based on industry research [9], and an average party size of 2.5 was used given the event's urban setting." This transparency builds trust in the results and highlights areas where actual data could improve future estimates.

Sensitivity Analysis and Risk Management

Estimating economic impact with incomplete data introduces uncertainty. Our model incorporates **sensitivity analysis** capabilities to handle this risk:

- **Parameter Sensitivity:** The user or analyst can define high/low values for key proxy inputs (e.g. ±20% on average daily spending, or using a range of travel party sizes) and re-run the model to see how outcomes change. The model could automate this by generating a **sensitivity table** or scenario analysis – for example, showing total impact under a “conservative scenario” (lower spending, shorter stays) vs. an “optimistic scenario” (higher spending, longer stays). This can be included in the report or provided as supplemental output. If proxies are used, we will often provide a **margin of error** or range. For instance, “Total impact is approximately \$22.2M (could range from ~$20M to $24M based on spending variability).”

- **Benchmark Validation:** When using proxies, the model draws from multiple reference points to ensure they are reasonable. For example, if no travel party data is available, the model might consider averages from several similar events (marathons in comparable cities) rather than a single source. This mitigates the risk of an outlier study skewing our assumptions. The IBRC method encourages using *prior surveys of similar events* to inform unknown inputs [14] [9], which we follow. We also allow the user to override any default assumption with their own input if they have specific knowledge (for instance, if they know anecdotally that many runners treated the event as a vacation, they might increase the assumed average nights).

- **Avoiding Double Counting:** The model’s logic is careful to avoid counting any spending twice or including spending that isn’t truly incremental. For example, local spectators’ spending on race day is not automatically counted unless there’s a justification that it’s additional (perhaps prevented from leaking out of the economy) [15]. This conservative approach reduces the risk of overestimation, which is a common criticism of impact studies. The user can choose to include or exclude certain components (like local organizer spend or local spectators) depending on their definition of impact, and the report will clarify that choice.

- **Use of Confidence Intervals:** If survey data is used, the model can convey statistical confidence. For example, with a large sample size, one might say the estimates are accurate within ±3% [39]. If proxies are used, the model might not have formal confidence intervals, but it will still highlight which inputs are estimates and encourage treating the results as approximations (e.g. rounding to the nearest \$10,000 or \$0.1M rather than giving a false sense of precision).

By handling uncertainties in this manner, the model remains robust. It provides not just a single-point estimate but context about reliability. In practice, this means decision-makers see a likely range of impact and can better gauge risk (for instance, understanding that if the average spending per visitor was overestimated, the real impact could be on the lower end of the range).

Integration and Automation Considerations

Our economic impact model is built with automation in mind, to eventually integrate seamlessly with a user-facing interface or API service:

- **Python Implementation:** The model can be implemented as a Python module or set of functions. Using libraries like **pandas** for data manipulation (to ingest survey spreadsheets or CSVs) and perhaps **NumPy** for calculations allows efficient handling of input data. The logic described (actual vs. proxy, calculations) will be encoded in these functions. This code can be wrapped into a **CLI tool or an API** using a web framework (like Flask/FastAPI) for programmatic access. The use of Python ensures compatibility with various systems and ease of maintenance or enhancement.

- **Spreadsheet Upload Workflow:** For a UI that allows users to upload an Excel template, the model will parse that file to extract the inputs (Table 1 fields). Clear instructions will be provided in the template so the user knows how to input their data (e.g., a cell for total participants, a cell for average spend if known, etc.). Once uploaded, the backend runs the model and generates the outputs. The structured report could be generated as a **PDF or HTML** file. We can use templating engines (like Jinja2 for HTML, which can be converted to PDF) or report libraries to insert the calculated values into a pre-designed report format. The Shamrock report layout can be replicated in a template, where dynamic fields (like total participants, total impact numbers, etc.) are filled by the model's results.

- **API Integration:** For more automated pipelines (for example, a scenario where multiple events' data are sent via API to get impacts), the model will expose an API endpoint. A JSON payload following the schema in Table 1 can be POSTed, and the response will contain the JSON of results (Table 2 metrics). Optionally, the API could also return a link or binary of the formatted report if requested. This design means the model could be plugged into larger analytical systems or a web application without human intervention on each run.

- **User Data Uploads:** In future iterations, we plan to allow more **user-generated data integration**. For example, an organizer could upload raw survey data (CSV of responses). The model would include a parsing component to map common survey questions to the needed inputs. (The appendix of the Shamrock report lists typical questions [40] [41]; if the user's survey follows a similar structure, the model can automatically compute, say, average party size from questions about how many people traveled with the respondent, etc.). For robustness, we'll include a configuration to let the user specify how to interpret their data columns if they differ from our defaults. Over time, as more surveys are processed, the system could even learn or refine the proxy defaults by averaging the results of multiple events.

- **Modularity and Maintainability:** Each component of the model (data input, direct impact calc, multiplier application, report generation) is modular. This makes it easy to update one part without affecting others. For instance, if a new study comes out with updated average spending figures or multipliers for sports events, we can update the proxy data library without changing the core logic. Similarly, if a new category of impact is considered (say, environmental impact or something outside economic scope), it could be added as an extension.

- **Testing and Verification:** The model would be tested against known economic impact reports (such as the Shamrock Marathon 2024 data or Grandma's Marathon study) to ensure it reproduces similar figures when given the same inputs. This will validate that our calculations align with industry-standard methods. Discrepancies can be analyzed to further refine proxy assumptions.

- **Security and Data Handling:** When integrating into a UI/UX, careful handling of user-uploaded data is important. Our design assumes data is not personally identifiable beyond aggregate event info. Survey data, if uploaded, should be handled securely and possibly anonymized when summarized (the model really only needs aggregated metrics from it). API endpoints will include validation to ensure inputs make sense (e.g., negative or extremely large values can be flagged).

- **Future Extensions:** The automated pipeline could eventually incorporate **real-time data**. For example, if during an event, live registration or hotel pickup data is available, the model could adjust some estimates on the fly. Also, as more events are analyzed, the system could offer benchmarking – allowing a user to see how their event's impact compares to similar events' impacts stored in a database.

In summary, the model is designed not just for one-off analysis but as a **reusable tool**. Its schema and logic allow it to be plugged into different interfaces (spreadsheets, web forms, APIs), making economic impact analysis faster and more accessible for endurance events of all sizes. By formalizing the process in code and clearly defined inputs/outputs, we reduce manual work and potential errors, while providing a consistent structure that users can trust.

Conclusion

The proposed Python-based economic impact model for endurance running events provides a **comprehensive yet flexible framework** to estimate how much these events contribute economically to host communities. By intelligently using available survey data and reliable proxies, the model adapts to data-rich and data-sparse situations alike. It covers all critical components – from direct visitor spending on hotels and meals to the broader ripple effects captured via multipliers – and presents the results in both a polished report format and structured data outputs. Key considerations like only counting non-local expenditures [1], using proven formulas [33], and applying local tax rates [42] ensure the estimates are credible and tailored to the event context.

Through robust design features (input schema, proxy logic, sensitivity analysis) and integration capabilities, this model can be a powerful tool for race organizers, city officials, and researchers. It not only quantifies the economic boost from events such as marathons, but also provides transparency into how the numbers are derived and how reliable they are. As a next step, this model can be implemented and iteratively improved with real event data, ultimately becoming an automated service that turns user inputs into actionable economic insight – helping demonstrate the value of endurance events to local economies and informing decisions on event planning and resource allocation.

---

[1] [2] [10] [12] [13] [20] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [35] [36] [37] [38] [39] [40] [41] [42] 2024 Shamrock EI Report.pdf
file://file-RGtNRNxL7tjd8ybTnGz3VY

3 9 14 18 19 33 34 Inexpensively Estimating the Economic Impact of Sports Tourism Programs in Small American Cities
https://www.ibrc.indiana.edu/ibr/2015/spring/article1.html

4 7 11 Economic Impact of Grandma's Marathon Weekend Reaches Nearly $40 Million, Study Shows - Running USA
https://www.runningusa.org/industry-news/economic-impact-of-grandmas-marathon-weekend-reaches-nearly-40-million-study-shows/

5 6 8 15 16 17 21 22 sportsdestinations.com
https://www.sportsdestinations.com/management/economics/blue-ridge-marathon-economic-impact-study-11885