P1

Morning City officials! 我们是 Lumberjack Consultancy。

P2+3

(Chapter 1)

费城的房产税收关乎城市的公平与未来。我们今天的目标非常明确：**帮助费城建立一个更准确、更公平的房屋估价系统，确保每一位市民的税负都公正反映其房产的真实市场价值。我们的研究方法可能与所有人不同，所以请尝试理解。**

Property taxes in Philadelphia are about fairness and the city’s future. Our goal today is clear — to help the city build **a more accurate and equitable** property valuation system. Our approach might be a bit different from others, but we will try our best to explain.

P4

Our Primary Dataset is **Philadelphia Property Sales** from Philly’s government website. We also used **Census ACS data** and **these datasets from OpenDataPhilly** as supplementary.

P5+P6

(Chapter 2)

**我们将2023年-2024年的房产价格绘制在地图上时，模式清晰可见：**

1. **不同等级的Residential sale price呈现出空间聚集特征**
2. **价格在空间上不是连续过渡，xxxx**

**从价格分布我们可以看出，高价房的的数值非常惊人的高，显示出了右侧长长的“尾巴”。**

**总结而言，费城的房价故事是一个高度“本地化”的故事。估价不能再只是简单地看房屋大小，它必须能理解这个房产所在的“小市场”。**

When we mapped 2023–2024 houses sale prices, the pattern jumped out:

prices cluster by area, and they don’t shift smoothly — they jump from block to block.  
The distribution also has **a long right tail** — those few luxury homes \*\*really\*\* stretch the scale.  
So the story of housing prices in Philly is a local one. A fair valuation has to **go beyond square footage** — it has to understand the **neighborhood market** it sits in.

P7

到底是什么原因造成了那些房子卖出高价，或者只能低价卖出呢？我们运用房地产估价中被的“比较销售法”统计周边历史成交房价，这反映了房产所处的局部市场环境，同时邻近的成交价直接影响买家和卖家的心理预期，构成直接的参照点。发现有很大的相似之处。

What makes one home sell high and another sell low?  
A big factor is called the “**comparable sales”** — looking at what similar nearby homes sold for.  
These past sales **define the local market** and **shape everyone’s expectations.**  
That’s why homes in the same area often **end up with similar prices.**

P8

在开始构建预测模型之前，我们必须强调本公司在数据处理上与其他机构**截然不同的核心哲学**：传统的做法是简单地**删除 (delete)** 销售价格（sale price）中的异常值（outlier）。然而，我们坚信保留所有数据告诉我们的信息，因为这些“异常值”往往反映了市场中可能被直接忽视的复杂或非主流现象，**保证评估的公平性**。

我们不会删除数据，而是1.我们**识别**出处于销售价格两端的极端值（Outliers），将其定义为不可靠数据；2.根据market price，**调整并替换**这些不可靠的销售价格。但是我们为了数据可靠程度，将那些两端的极端值定义为不可靠数据，在后续分析中赋以较低的权重。

我们的方法是在**保留市场信息的完整性**的前提下，通过**专业、量化的调整手段**，确保用于模型训练的数据是可靠且公平的。

Before building our prediction model, we want to **highlight a key difference** in our approach to data: Most teams delete “outliers.” We don’t.

we believe that every data point tells a story.

These so-called “outliers” often capture **unusual but meaningful market behavior**s — removing them could **erase important signals** about equity and market dynamics.

So our first step is that Identifying **extreme sale prices at both ends** as” less reliable”;

Next, we **adjust these values** based on their relationship to the market price, \*\*in the mean while\*\* **assigning them lower weights** in later analyses.

This way, we keep the data honest — and the model fair.

P9

为了抓住这个“本地化”的故事，我们采取了“层层递进”的模型构建策略。如同盖摩天大楼，我们从地基开始，逐步增加模型的复杂性和精度。我们发现，每增加一层关于“社区环境”的变量，模型的预测准确性就显著提高。最终模型将误差的平均大小从xxx降低到了xxx，这在实际货币价值上代表了对基线模型 **超过xxx%的改进！**

To tell this local story right, we built the model **layer by layer** — just like building a skyscraper.  
Every new layer of data made our predictions **sharper and more accurate**.  
By the end, our model cut the average error from XXX to XXX, an improvement of more than XXX% over the baseline — real progress toward fairness and accuracy.

P10

在对模型的验证中，我们的RMSE值在四次模型的逐步改进中逐渐下降，这意味着我们的模型计算误差从最开始的xxxx减少到了xxxx，预测模型可靠性得到了显著提升。这是在最大可能维持预测公平性与普适性的情况下，得到的非常稳定的预测结果。

As we refined the model step by step, the RMSE, which means the model’s prediction error, kept going down — from XXXX to XXXX.  
That drop means our predictions became much more reliable.

And what matters most is that **we achieved this improvement** without sacrificing fairness or consistency — the model **stayed stable and balanced** eventually tho.

P11+P12

(Chapter 3)

我们提炼出了三大价格驱动力：

1.宜居面积 (Livable Area)： 这是传统因素中的最强者，证明了结构性特征的重要性依然不容忽视。

2.周边历史成交房价 (Comparable Sales)： 模型创新性地引入了周边房产的实际售价。这在房地产估价中被称为“比较销售法”，它让我们的模型能够像一个经验丰富的估价师一样，使用最新的市场参照点来校准价格。

3. 邻里固定效应 (Zip Code Fixed Effects)： 这是最重要的因素。 它捕捉了特定邮政编码内，所有其他因素无法解释的、纯粹的“地段价值”和市场情绪。

We found three big forces behind housing prices:

1. Livable Area – The **stronges**t among traditional factors, proving that structure of houses still matters.
2. Comparable Sales — nearby sale prices help the model **think like a real appraiser,** using fresh market references.
3. Zip Code Fixed Effects — this is the game-changer; it captures the **true “location value”** and **local market mood** that numbers alone can’t explain.

P13

通过分析预测值与实际成交价的差距我们发现：

我们的模型系统性地高估了xxxx（zipcode/census tract），这些地区往往是(**收入较低、房龄较老)**的社区。这意味着，在当前系统中，这些社区的居民可能面临相对不公平的高税收评估。而我们观察到xxxx（zipcode/census tract）存在系统性低估。这些社区通常是（**快速绅士化 Gentrifying)或靠近新开发项目）**的地区。这表明市场正在快速行动，但估价系统未能及时跟上。

When we looked at the gap between predicted and actual prices, a pattern stood out:  
Our model tends to **overestimate homes in lower-income, older neighborhoods** — meaning those residents may be **paying more than their fair share** in property taxes.  
Meanwhile, it **underestimates homes in fast-changing areas** near new developments — showing that the market is moving faster than the city’s valuation system can keep up.

P14

当然我们的数据本身也存在一定偏差，这种偏差也会影响预测的公平性：

富裕地区的数据覆盖有限，这些高价值住宅过于独特，模型难以泛化。

同时，富裕社区的记录往往更完整，而贫困社区数据缺失，造成模型在空间表现上的不均衡。

此外，还存在价格层面的偏差：高价房的成交价往往低于挂牌价，而低价房更接近挂牌价。  
这导致模型系统性地高估高价房、低估低价房，从而带来潜在的公平性问题。

Of course, our dataset themselves **contains certain biases,** which also affect the fairness of our predictions.  
Records in **wealthier neighborhoods** tend to be more complete, while poorer communities suffer from data gaps — creating **spatial imbalance** in model performance.

P15

基于这些发现，我们提出两项核心政策建议：

1. **立即校准系统：** 利用我们的模型结果，立即对那些系统性高估的低收入社区进行**评估值修正**，确保税负的公平分配。
2. **整合空间数据：** 建议OPA**常态化整合**我们发现的有效空间特征，尤其是 **周边成交价** 和 **邻里固定效应**，进入下一代 AVM，以提高对快速变化市场的响应能力。

Based on these findings, we propose **two key policy recommendations:**

For example, We suggest that OPA regularly **include spatial features**—like nearby sale prices and neighborhood fixed effects—in future AVMs.

That way, the system can **keep up with fast-changing markets** and stay fair over time.

P16

感谢您的时间。我们深信 Lumberjack Consultancy 的方案能为费城的财政公平之路提供清晰、可靠的指引。我们现在很乐意回答您的问题。

Thank you for your time.  
We truly believe that our proposal can offer Philadelphia a **clear and reliable path** toward fiscal fairness.  
We’d now be happy to take your questions.

（用周边房价作为自变量的思路，在房地产估价中被称为\*\*“比较销售法”（Comparable Sales Approach），而在统计模型中，这属于利用空间自相关性（Spatial Autocorrelation）和邻近效应（Neighborhood Effects）\*\*。周边房价反映了房产所处的局部市场环境，这包括：社区吸引力：好的学校、低犯罪率、便利设施（公园、商店、交通）。土地价值：同一地理区域内的土地价值通常相似。市场情绪：邻近的成交价直接影响买家和卖家的心理预期，构成直接的参照点。）