

基于SAS的轨迹数据可视化分析

王有芹

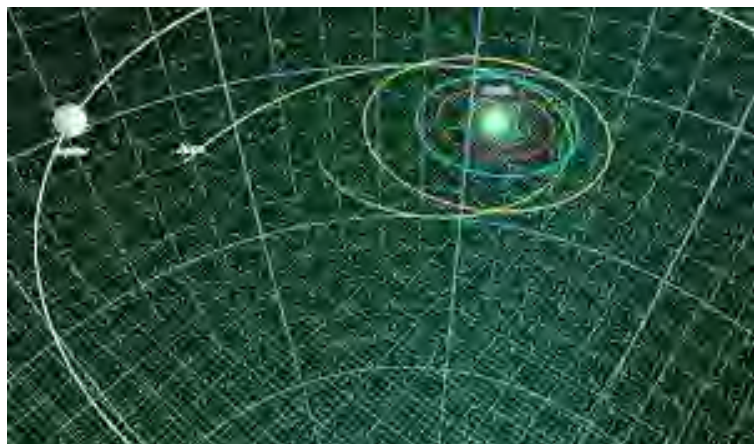
SAS Research & Development (Beijing) Co. Ltd

Email: youqin.wang@sas.com Wechat: cindy10082

提纲

- 时空轨迹数据概述
- 轨迹数据分析的研究综述
- 案例
 - 案例背景
 - SAS分析工具简介
 - 轨迹数据的可视化分析

时空轨迹数据概述



时空轨迹数据概述



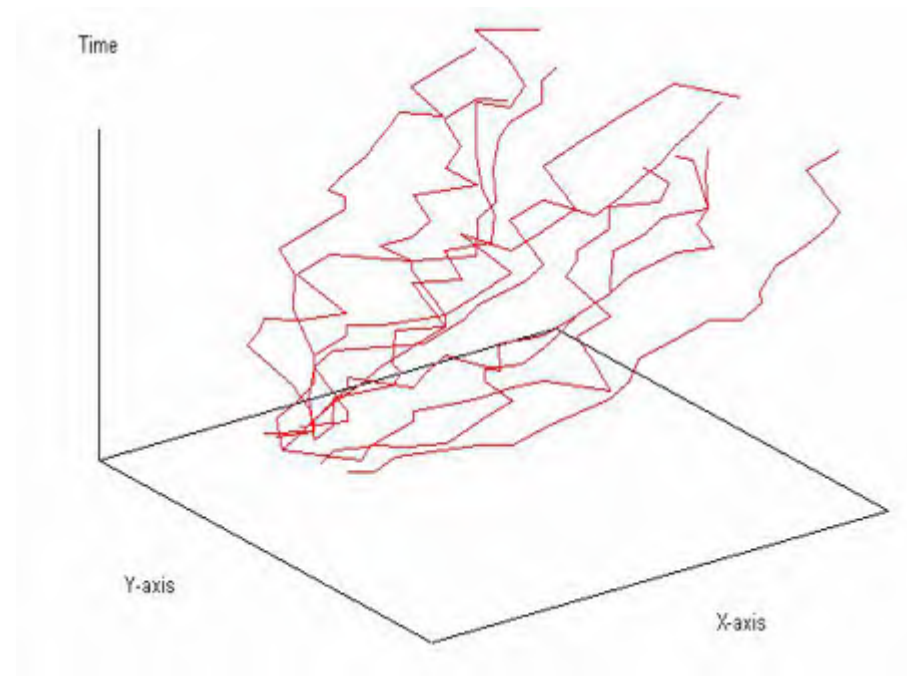
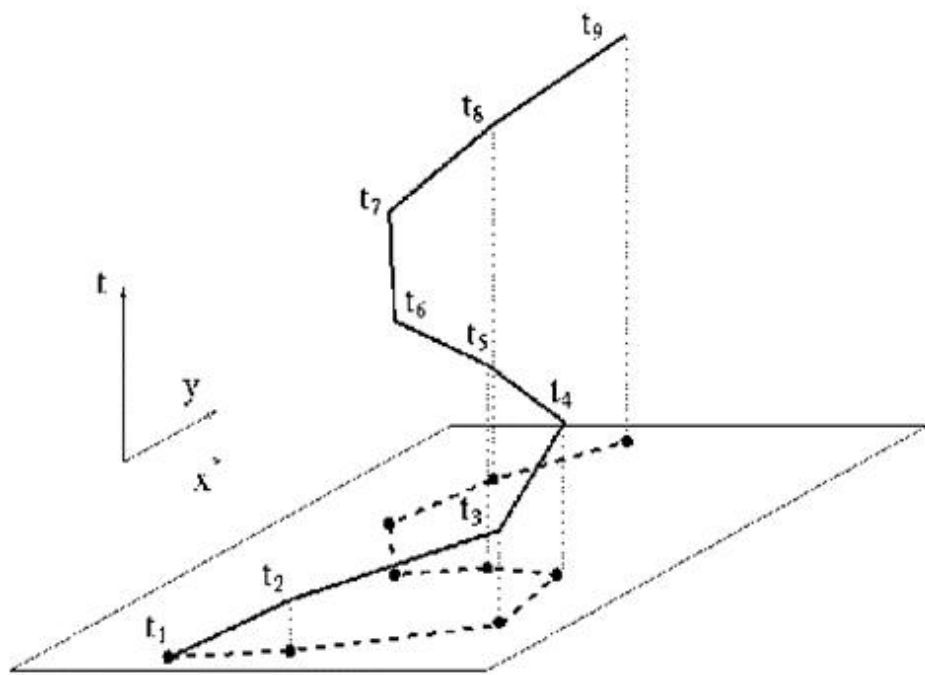
A **trajectory** is the path that a moving object follows through space as a function of time.

Tid	geometry		timest
1	48.890018	2.246100	08:25
1	48.890018	2.246100	08:26
...	...	...	...
1	48.890020	2.246102	08:40
1	48.888880	2.248208	08:41
1	48.885732	2.255031	08:42
...	...	...	...
1	48.858434	2.336105	09:04
1	48.853611	2.349190	09:05
...	...	...	...
1	48.853610	2.349205	09:40
1	48.860515	2.349018	09:41
...	...	...	...
1	48.861112	2.334167	10:00
1	48.861531	2.336018	10:01
1	48.861530	2.336020	10:02
...	...	...	...
2	...	...	...

时空轨迹数据概述

- **Trajectory Data**

- **Represented by a set of location and time data (Spatio-Temporal)**



时空轨迹数据概述

- Trajectory Data Representation:

- Geometry-based

$T=(x_1,y_1,t_1), \dots, (x_n,y_n,t_n) \Rightarrow$ position in space (x_i,y_i) at time t_i

- 2D , 3D

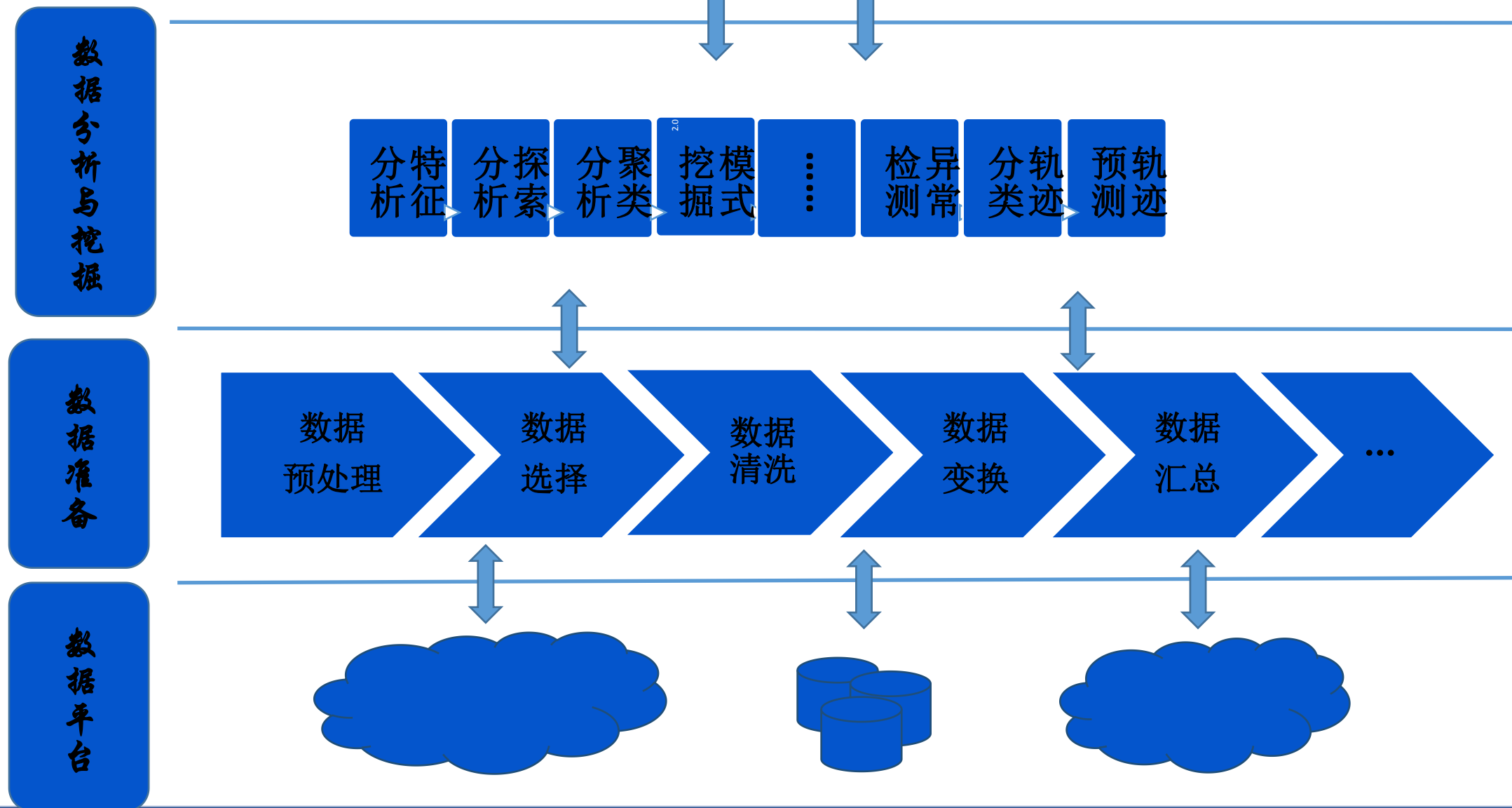
- Semantic-based

$T=(p_1,t_1), \dots, (p_n,t_n) \Rightarrow$ position point p_i at time t_i

- Trajectory Data Characteristics

- *Volume + Velocity + Variety + Value*

轨迹数据的业务应用场景

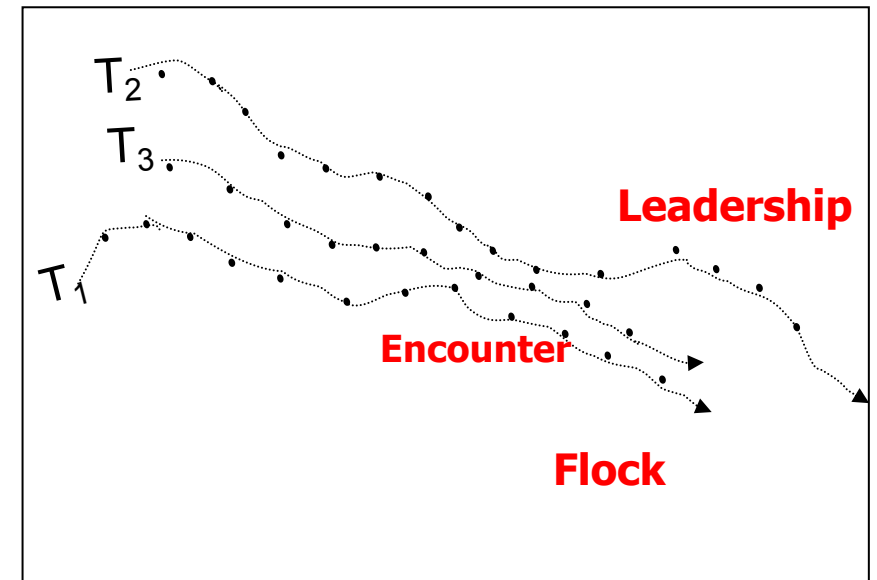


Trajectory Data Analysis

- **Spatial perspective**
 - Position - Coordinates; Points
 - Distance, Direction, Spatial Extent
 - Distribution
- **Temporal perspective**
 - Time instance, Time interval, Duration
 - Distribution
- **Spatio-temporal**
 - Speed/Velocity
 - Acceleration Rate
 - Distribution

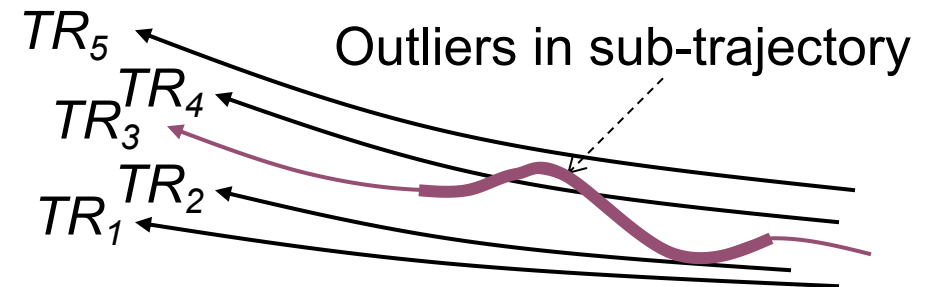
Trajectory Data Mining

- Trajectory Patterns
 - Relative Motion Patterns
 - Flock pattern, Leadership pattern, Encounter pattern
 - Frequent Trajectory Patterns
 - Group pattern, Co-location pattern
 - Trajectory Sequential Patterns
 - T-Patterns,
 - Periodical Patterns
 - ...



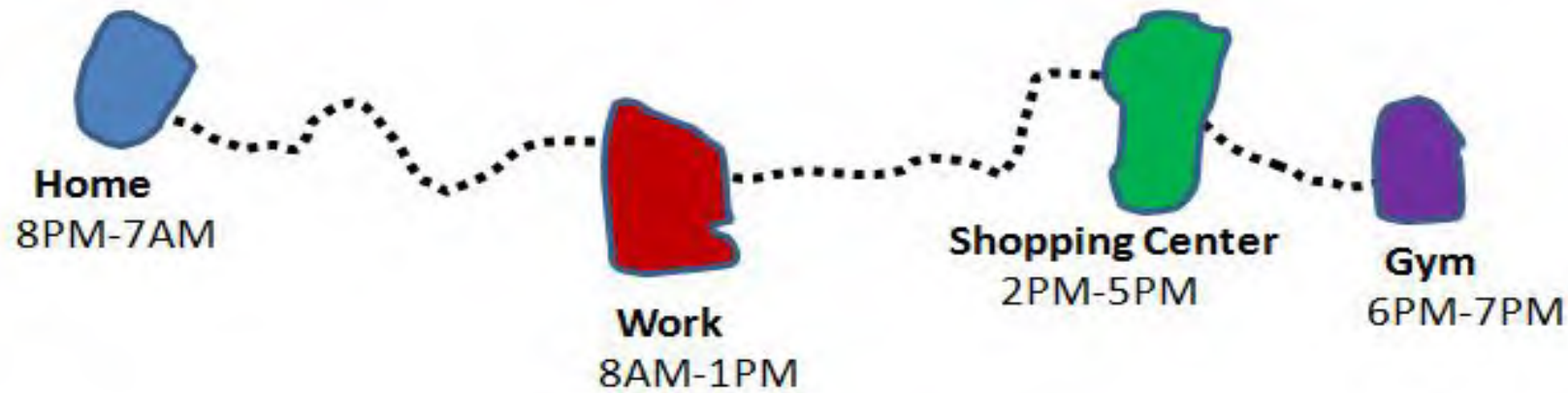
Trajectory Data Mining

- Trajectory Clustering
 - Region-based, trajectory-based (sub-trajectories)
 - Density, Distance
- Trajectory Outlier Detection
- ...



Trajectory Data Mining

- Semantic-based trajectory data mining:
 - Patterns are computed based on semantics of the data
 - Generally, trajectories are pre-processed to enrich the data



轨迹数据分析的研究案例

• 案例背景

- 一天的轨迹数据：包含了大约535万条记录
- 园区配备的传感器信标进行数据采集
- 轨迹数据集的数据格式为：

时间戳，游客ID，活动类型，X坐标，Y坐标

轨迹数据集的样例数据如下：

2014-6-06 08:00:55, 483896, check-in, 63, 99
 2014-6-06 08:00:55, 539617, check-in, 0, 67
 2014-6-06 08:00:55, 932874, check-in, 0, 67
 2014-6-06 08:00:59, 179386, movement, 63, 98
 2014-6-06 08:00:59, 1053224, check-in, 99, 77
 2014-6-06 08:00:59, 580064, check-in, 99, 77
 2014-6-06 08:01:04, 179386, movement, 64, 97

.....



轨迹数据分析的研究案例

• 分析工具

- SAS Enterprise Miner (SAS EM)
 - SEMMA
- SAS Visual Analytics (SAS VA)
 - High-performance
 - BI, Exploration, Analysis
 - Self-service
 - Interactive
 - In-memory



轨迹数据分析的研究案例

- 数据准备

- 侧重语义特性的分析
- 数据映射：坐标数据 => 景点
- 数据清洗、变换
- 数据转置、分拆
- 数据选择、汇总
- 数据排序、分层
- 。 。 。

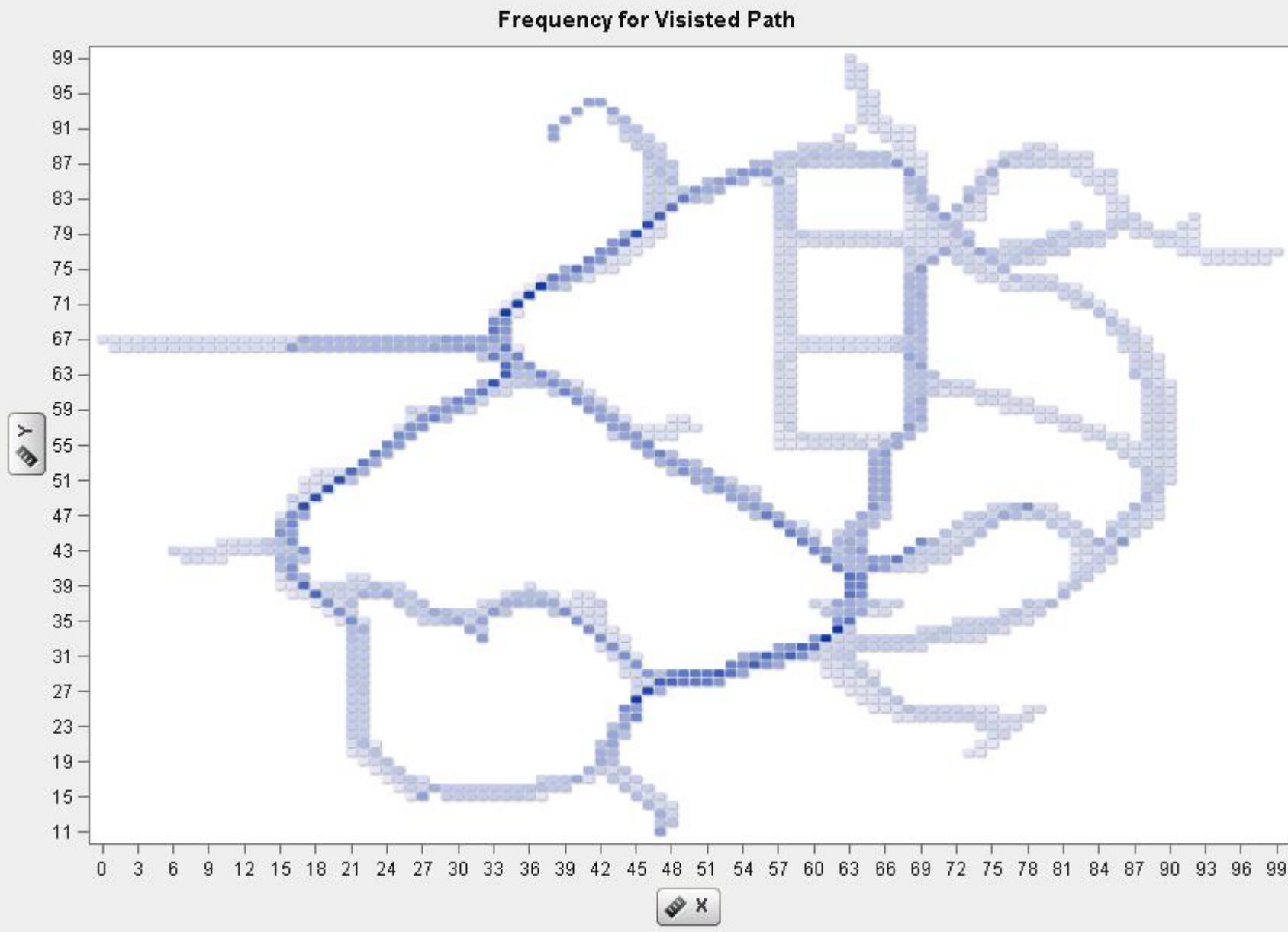
Attraction ID	Attraction Name	Attraction Type	Attraction Typeid	X	Y
1	Wrightiraptor Mountain	Thrill Rides	1	47	11
2	Galactosaurus Rage	Thrill Rides	1	27	15
3	Auvilotops Express	Thrill Rides	1	38	90
4	TerrorSaur	Thrill Rides	1	78	48
5	Wendisaurus Chase	Thrill Rides	1	16	66
6	Keimosaurus Big Spin	Thrill Rides	1	86	44
7	Firefall	Thrill Rides	1	17	43
8	Atmosfear	Thrill Rides	1	45	24
	...				
69	West Entrance	Entrance	10	0	67
70	North Entrance	Entrance	10	63	99
71	East Entrance	Entrance	10	99	77

轨迹数据分析的研究案例

- 数据分析和挖掘
 - 探索分析
 - 轨迹热图
 - 访客量分析
 - POI 特征分析
 - 最热门景点的排名
 - 游客停留时长的波动
 - 分时段客流量分析
 - 访客行为分析
- 路径分析
- 频繁模式
- 聚类分析
- 异常值分析

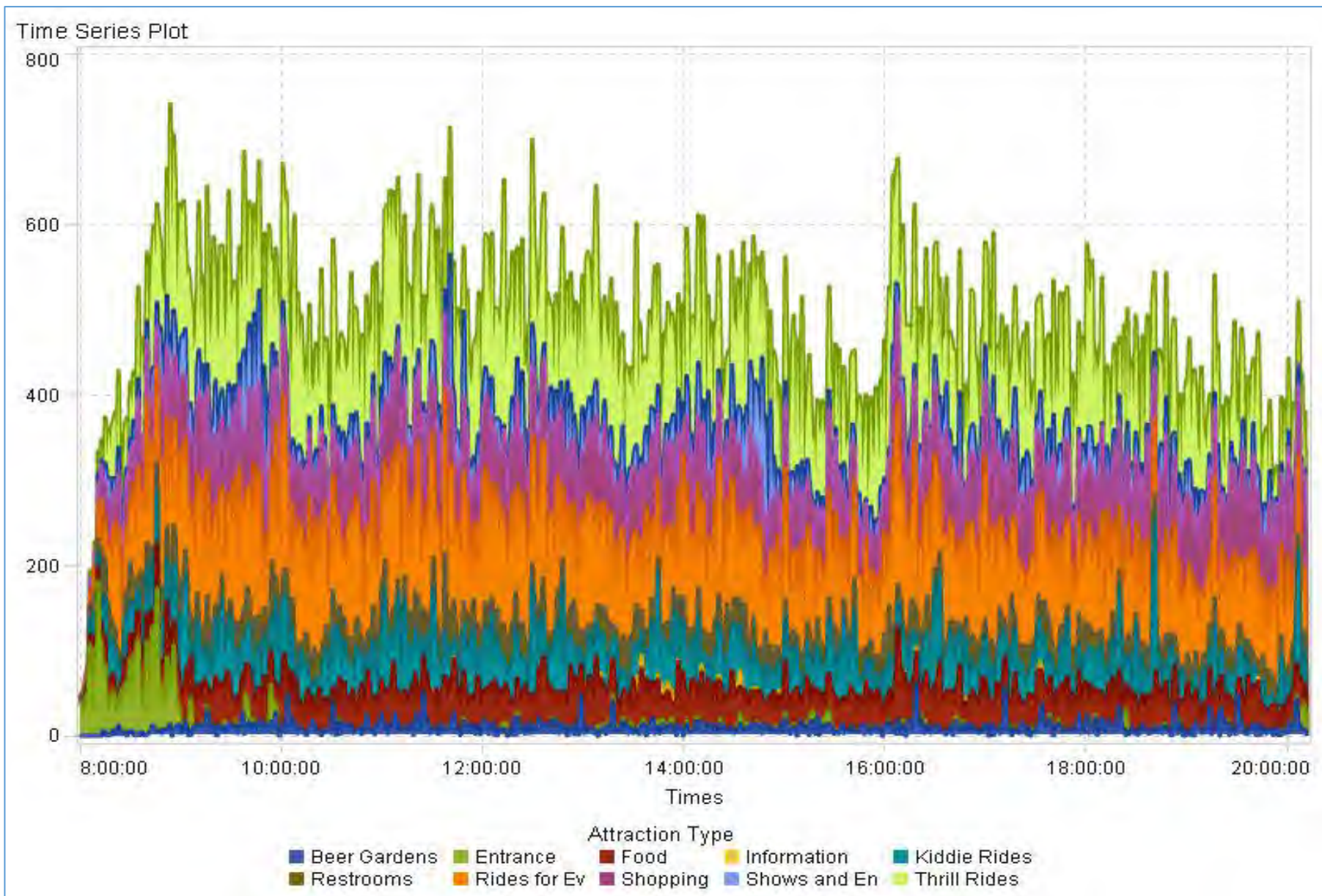
数据分析

- 探索分析
 - 轨迹热图



数据分析

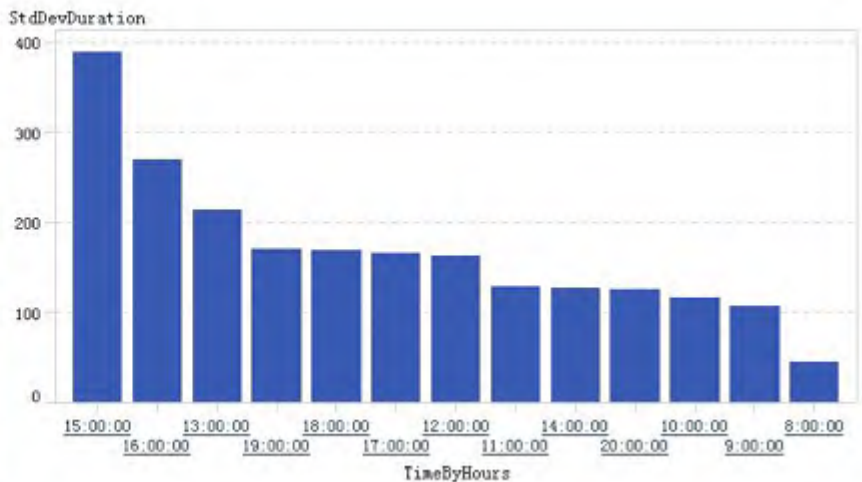
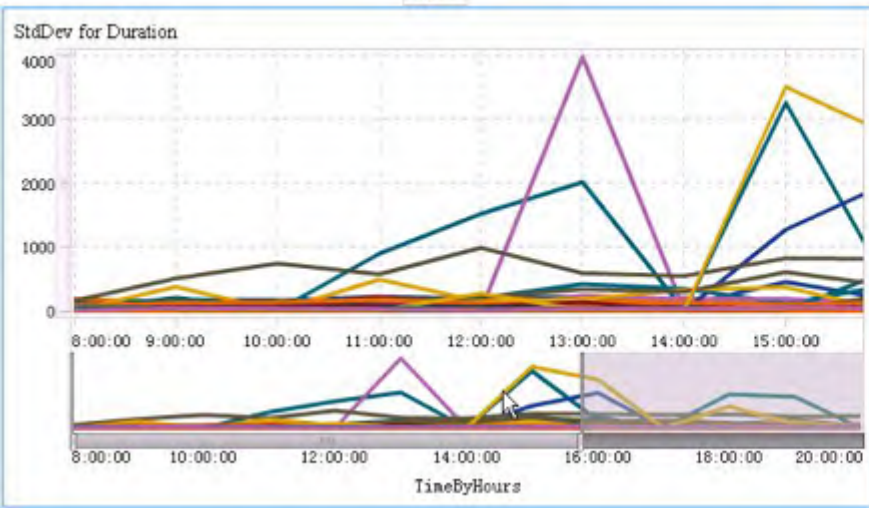
- 探索分析
 - 访客量
 - 分时段
 - 分类型



数据分析

- POI特征分析
 - 游客停留时长的波动

分时景点访问量Top10 景点访问量 Top10 景点停留时长 Section 3			
attractionname	attraction areal	Frequency	Duration
Alvarez Beer Garden	Bear Gardens	4,836	253113
Atmosfear	Thrill Rides	12,456	193350
Auviletops Express	Thrill Rides	8,377	85829
Beelzebubo	Kiddie Rides	1,787	13336
Blue Iguanodon	Kiddie Rides	1,690	15586
Chenrational Sweets	Food	648	4991
Creighton Pavilion	Shows and En	9,291	266149
Cyndisaurus Asteroid	Kiddie Rides	2,793	29036
Daily Salab Maps and In	Information	1	16
Darwin's Stop	Restrooms	4,219	30885
Dino Chic Clothing	Shopping	2,704	21781
Dykesadactyl Thrill	Rides for Ev	5,687	41131
East Entrance	Entrance	1,074	433427
Eberlesaurus Roundup	Rides for Ev	14,027	107637
EberTrex Fries	Food	6,136	45785
Enchanted Toadstools	Kiddie Rides	2,270	16461



数据分析

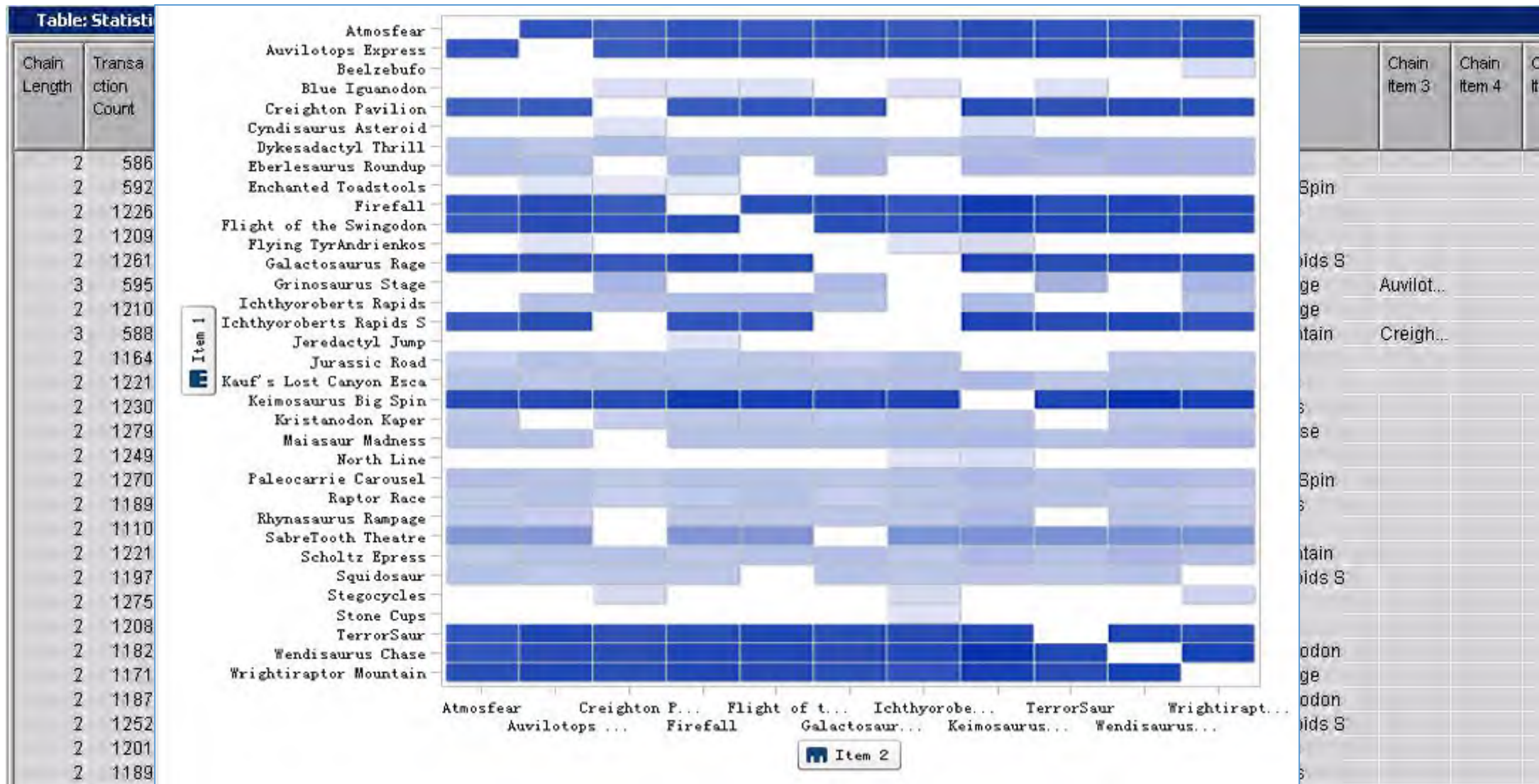
- POI特征分析
 - 访客行为分析
 - 入口数据



数据挖掘——路径分析

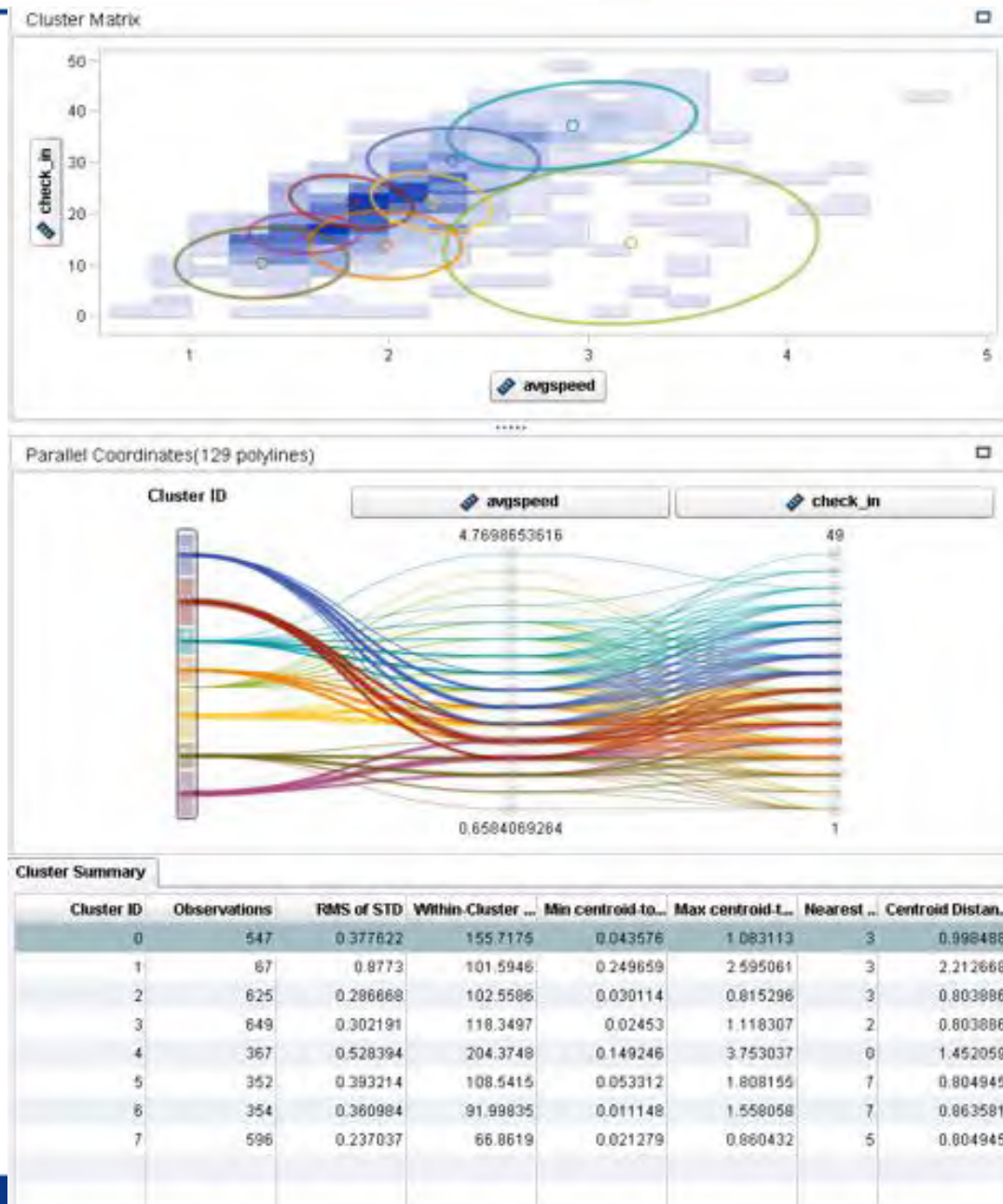
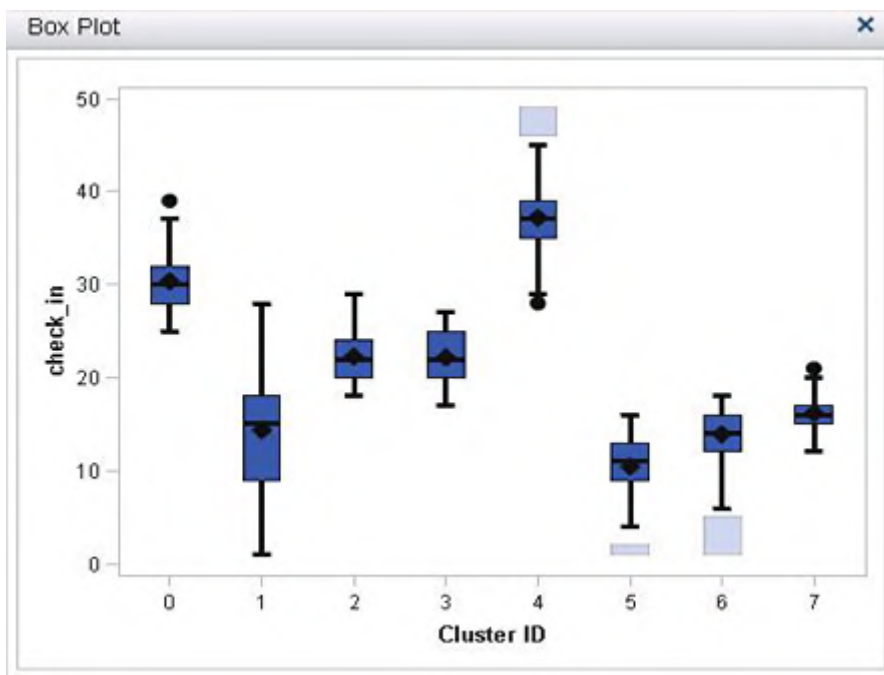


数据挖掘——频繁模式

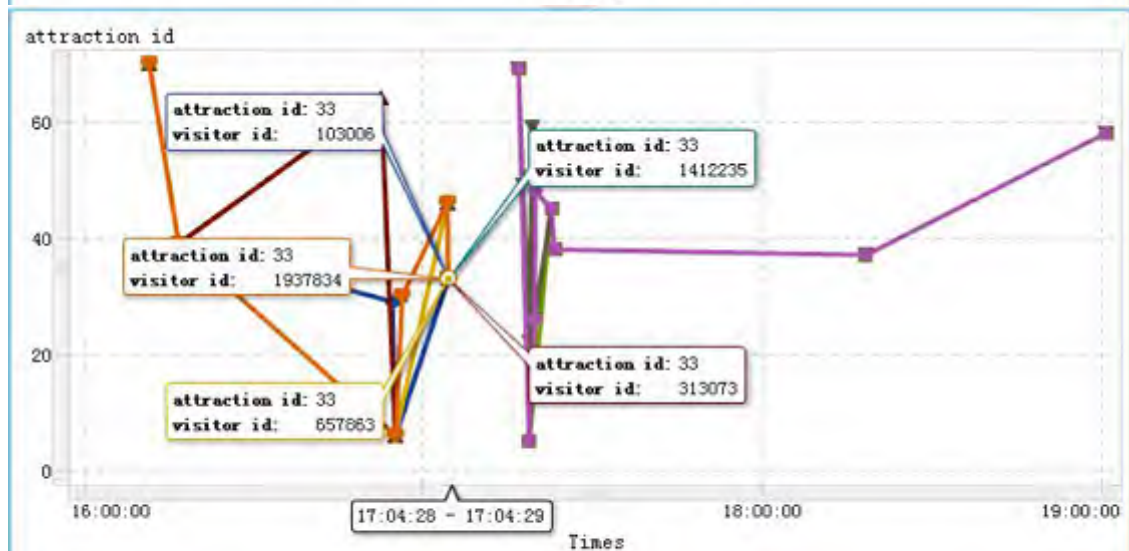
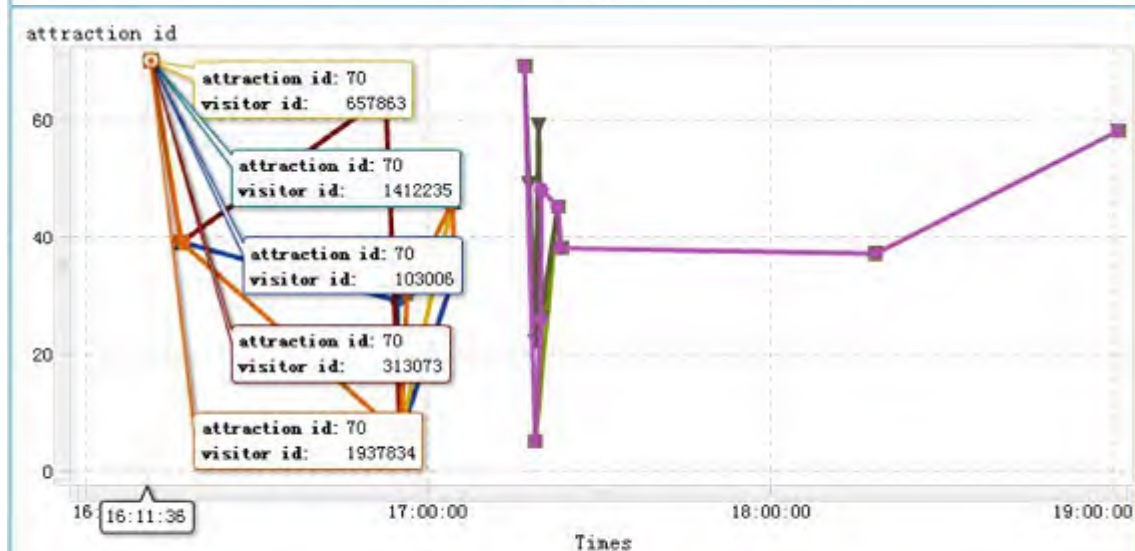
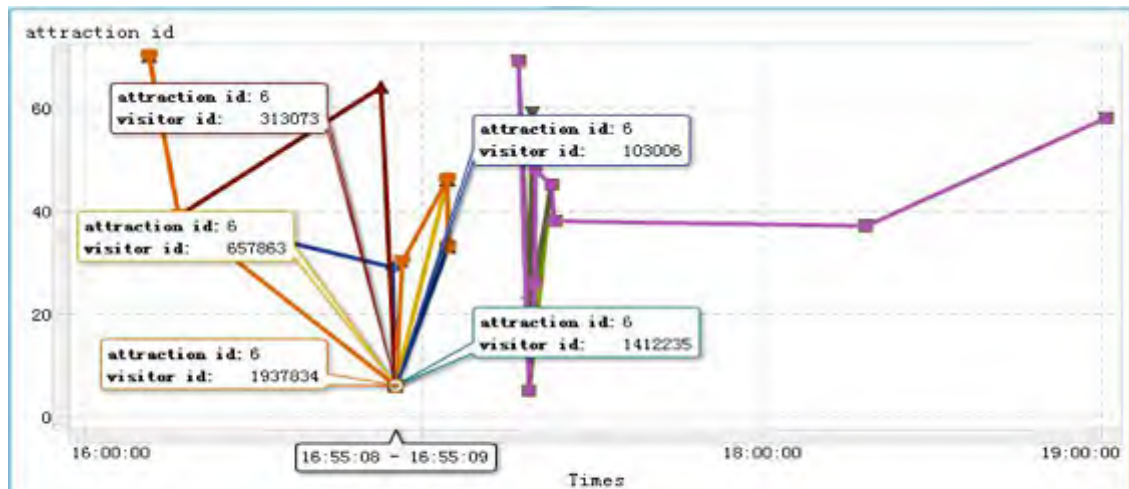
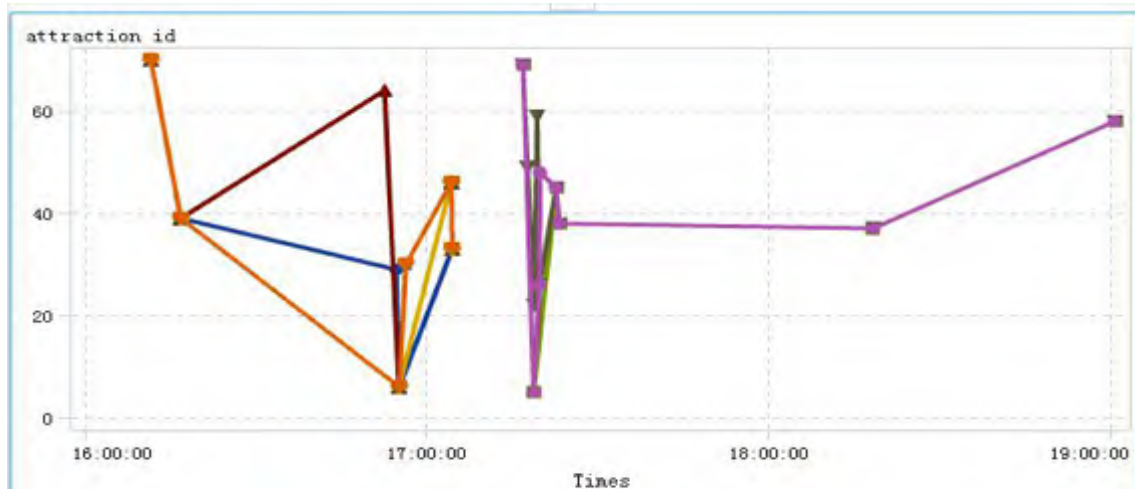


数据挖掘——聚类分析

- 聚类
 - Average speed
 - Check-ins



数据挖掘——异常值分析



visitor id

103006 195087 313073 657863

1412235 1614660 1937834 2096715

visitor id

103006 195087 313073 657863

1412235 1614660 1937834 2096715

Future



谢谢大家！